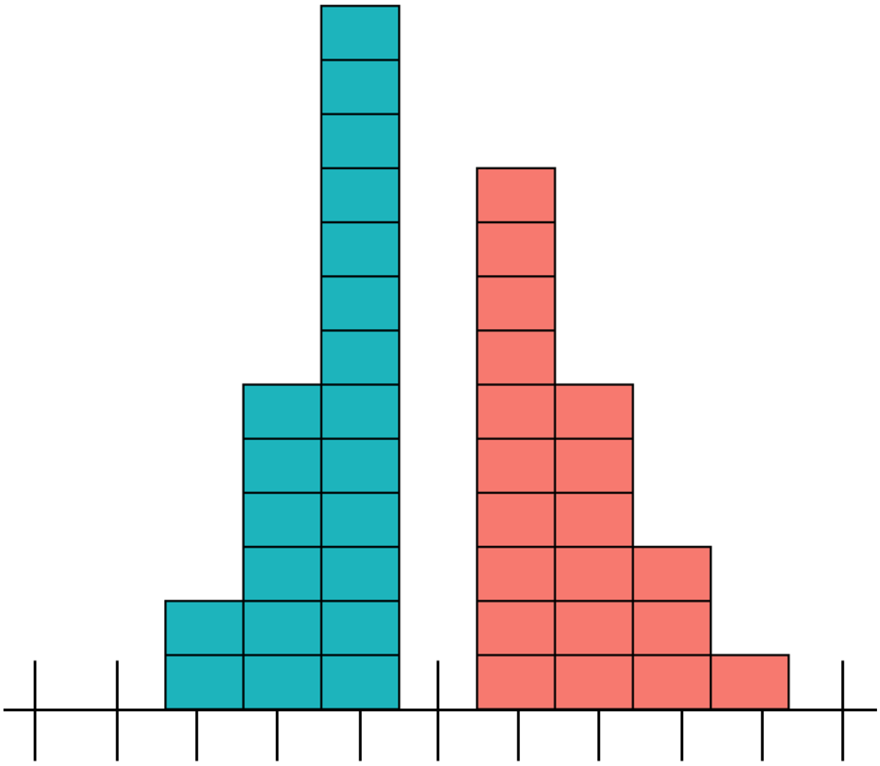
CHALMERS
UNIVERSITY OF TECHNOLOGY



Fill or kill limit order costs in currency markets

An empirical pricing model

Master's thesis in Engineering Mathematics and Computational Science

Niklas Forsström

Department of Mathematical Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2020

Master's thesis 2020

Fill or kill limit order costs in currency markets

An empirical pricing model

NIKLAS FORSSTRÖM

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Mathematical Sciences
Division of Applied Mathematics and Statistics
Chalmers University of Technology
Gothenburg, Sweden 2020

Fill or kill limit order costs in currency markets
An empirical pricing model
NIKLAS FORSSTRÖM

Supervisor: Torbjörn Hovmark, Ampfield AB
Examiner: Simone Calogero, Department of Mathematical Sciences, Chalmers University of Technology

Master's Thesis 2020
Department of Mathematical Sciences
Division of Applied Mathematics and Statistics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Illustration of a limit order book.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2020

Fill or kill limit order costs in currency markets
An empirical pricing model
NIKLAS FORSSTRÖM
Department of Mathematical Sciences
Chalmers University of Technology

Abstract

Fill or kill limit orders offer a way to execute an order, whilst imposing constraints on the execution price. They are therefore an attractive instrument for any speculating market participant. To make full use of the product, a model which describes the expected capital gain for a given set of market conditions is highly beneficial. Such a model can be leveraged in an algorithmic trading setting to make statistically sound market predictions.

This Master's thesis proposes and derives a model of the cost structure for fill or kill limit orders in currency markets. The order book is modeled by means of a Markov chain, which is used to predict the short term movements of the ask level. The underlying assets fair value is estimated based on the order book history. The market activity observed by a market participant is modeled as a function of the time of day, the latency, as well as the time since the last market event. A compound model based on these is evaluated and the model is found to offer significant improvements over a reference model. Further investigations of market activity modelling and market maker behaviour, particularly at times following events of high market impact, should be carried out.

Keywords: Limit Order, Empirical model, Futures, Order Book, Fill or Kill.

Acknowledgements

My sincerest gratitude goes out to Simone Calogero, who was my supervisor on behalf of Chalmers University of Technology. Without your flexibility and willingness to supervise this project at such an unconventional time, this paper would have never seen the light of day. I also want to thank the entire team at Ampfield for your meaningful discussions and valuable support.

Niklas Forsström, Stockholm, Oct 2020

Contents

List of Figures

1
Introduction

Ampfield AB is a Swedish proprietary trading firm based in Stockholm. The company invests in futures contracts in commodities and currency markets. The investment decisions are made by mathematical models, aiming to predict trends in the markets. This Master's thesis project was carried out on behalf of Ampfield AB.

1.1 Background: The rise of algorithmic trading

Algorithmic trading was first introduced in American equities markets during the 1990s. Since then, rule-based investing and the strive towards automated investment strategies has become ever more dominant and now serve as a common mode of trade.[1] These electronic markets vary in many ways - but they generally accept two types of orders, market orders and limit orders.

A market order is an order to purchase or sell a security at the best price possible, as soon as the order reaches the market. A market order ensures the buyer or seller that the order will execute, but it does not place any guarantees on the execution price. A limit order is a type of order to purchase or sell a security at a specified price or better. For buy limit orders, the order will be executed only at the limit price or a lower one. For sell limit orders, the order will be executed only at the limit price or a higher one. This stipulation allows traders to better control the prices at which they trade. While a limit on the order price is guaranteed, the filling of the order is not, and limit orders will not be executed unless the underlying price passes the limit order criteria. If the asset does not reach the specified price, the order is not filled and the investor may miss out on the trading opportunity.

If the order cannot be filled immediately it is normally queued in the order book. Each securities exchange uses its own specific algorithm, called a matching algorithm, to match orders. They broadly fall under two categories: first-in-first-out (FIFO) and pro-rata. Note however that exchanges such as cmegroup.com uses a mix of the following systems: Allocation, FIFO, FIFO with LMM, FIFO with Top Order and LMM, Pro-Rata, Configurable, Threshold Pro-Rata, Threshold Pro-Rata with LMM.

A basic FIFO algorithm, also known as a price-time-priority algorithm, works by giving the earliest active buy order at the highest price priority over any subsequent order at that price. These orders all take priority over any active buy order at a lower price. A basic pro-rata algorithm works by prioritizing active orders at a particular price proportional to the relative size of each order. Hence, all active orders at the same price will get filled in equal proportions upon the arrival of an opposing order.

1.2 Pricing of limit orders

When pricing a limit order, one can draw inspiration from options. A limit order does not fall under the category of options, but it has a similar structure and can therefore be priced in the same manner. To see why, consider short selling an American call option. If the buyer chooses to capitalize on the execution right, the seller is obliged to comply. If the buyer chooses to withhold from executing the option, the seller cannot affect this decision. Similar dynamics can be found after placing a limit order. Consider a trader who places a limit order in a financial market. Under some market conditions, the limit order will be executed and the trader will be obliged to trade. However, under different market conditions, the order might not get executed. In this case the trader will go without a trade. In both examples, the trader attains an obligation to execute, yet he does not have the authority to force execution.

The various queuing algorithms complicates the dynamics of execution and therein the pricing of the order. This problem can be alleviated by considering so called fill or kill orders. These orders either fill immediately upon arrival at the brokerage, or immediately get canceled. For a fill or kill order, the lifetime is essentially the same as the latency experienced by the trader, i.e. the delay between the order being placed and the order arriving at the order book. Since the latency for a given marker participant is fixed and the order will either get filled or killed at this predetermined time, this order shares commonalities with European options. Moreover, the order will only be executed if the market price is lower than the limit price. This generally causes the price of the order to be negative, which makes it more natural to consider the cost of placing such an order[1].

The latency varies between market participants and depends on factors such as geographical distance to the exchange, evaluation speeds of the trading algorithms etc. This latency is typically in the tenths of seconds, but some actors have a latency of less than a millisecond. The short latency renders standard methods, such as the Black and Scholes method, useless for pricing the limit order. Instead, market micro structure effects become predominant.

[1]There are scenarios in which the price could be positive. An example would be a market with low liquidity, where the asset has an obvious miss pricing. These scenarios are however unlikely.

1.3 Purpose

The thesis aims to empirically derive the cost structure for fill or kill limit orders, with enough accuracy to be used in algorithmic trading strategies. The model is primarily developed with high latency traders in mind. The purpose is to gain an understanding for the driving market factors behind the cost of placing such orders. Determining which of these factors play the most significant roles in affecting the cost, as well as the manner in which they do so.

1.4 Methods and working procedure

The project was initialized by a theoretical examination of the problem. Key assumptions such as the separability of the market activity from the order book evolution were formed and the quantities of interest were determined. Next, the market activities of futures contracts on Australian dollars (AD), British pounds (BP), Canadian dollars (CD) and Japanese yen (JY) were modeled. Models of varying complexity were evaluated and compared to strike a good bias–variance tradeoff.

After this came the modeling of ask level transition probabilities. The state of the order book was modeled by Markov chains, where the state representations varied in complexity. The order book data is discrete in time - and values are only recorded after a new market event is registered. Therefore, the models were initially compared on a stepwise basis to avoid the added complexities from the market activity modelling. Combinations from the two sets of models were then combined to form compound models. These could evaluate the transition probability as a function of time, rather than a function of the number of market events.

The next step was to evaluate different models of the fair asset value. There is no trivial way in which the fair asset value can be measured - and different models were used to predict the fair value based on the order book history. Lastly, this was combined with the models for ask level transition probabilities to form estimates of the cost for placing a fill or kill limit order. These final models were compared and evaluated. Due to the large datasets as well as the stringent performance constraints posed on the trading algorithms, the existing system was developed using C++.

1.5 Delimitations

This thesis is only concerned with the pricing of fill or kill type limit orders. Empirical investigation of persistent limit orders introduces additional complications related to biases in the data. Moreover, the thesis is only concerned with the pricing of buy orders. The extension to fill or kill sell limit orders can be done naturally. The contracts of interest have been restricted to currencies, but a similar analysis could easily be extended to commodities.

The cost is evaluated purely with the historical information contained in the order book. External information such as news, social media and press releases are not taken into account. The limit price is always assumed to be the same as the current ask price in the order book. Due to the short latency, it would be irrelevant to consider fill or kill limit orders with limit prices much different from the ask price at the time of placement. Moreover, the opportunity cost associated with not getting an order executed is neglected. One could easily include such a penalty term, which can be useful in the implementation of trading strategies. Lastly, with respect to confidentiality agreements, parameter values have been excluded from the report.

1.6 Structure of the report

The report starts with an outline of the underlying theory in chapter 2. In the chapter, more detailed explanations of futures contracts and order books are given. The chapter transitions into a technical discussion of how fill or kill limit orders can be priced - and what components are of key interest for the analysis. Methods of evaluating these quantities are discussed in detail. Chapter 3 describes the practical procedures associated with evaluating the quantities in chapter 2. It also describes the methods by which the different models were compared. The section contains figures, which illustrate the performance of the models and facilitate comparisons. Finally, chapter 4 contains a discussion of the methodology, as well as a conclusion on the acquired results. Recommendations for further investigations are given along with the final remarks.

2

Theory

This chapter gives an in-depth explanation of the theoretical assumptions and derivations used to evaluate the cost of placing a fill or kill buy limit order.

2.1 Forwards and futures contracts

A forward contract is an agreement between two parties to trade an underlying asset, for some pre-specified price K, at a fixed time T in the future. The forward price K is set such that the price of the forward contract equals zero at the time t of stipulation. Forward contracts were introduced as a way for companies to hedge the risks associated with their line of business. An example would be an airline company hedging the risk of a rise in oil prices. For this reason, forward contracts tend to be non-standardised and traded Over The Counter (OTC).

Futures contracts are standardized versions of forward contracts, which facilitate trading on an exchange. Much like forward contracts, futures contracts come with the obligation to purchase a security at a predetermined future price and date. Futures contracts are marked to the market on a daily basis. In other words, as futures prices change, daily cash flows are made - and the contract is rewritten in such a way that the value of futures contract at the end of each day remains zero.[2] Due to the contracts rewriting, the dividends at the end of each trading day may be negative. Therefore, an investor is normally obliged to maintain a safety margin. A futures contract can be closed at any time, without incurring a cost. This is practically achieved by assuming a short position in the same contract, causing a net zero effect.

Many futures contracts have a delivery date. If the position in the contract is not closed before this date, physical delivery of the underlying asset will occur. Since the majority of market participants are speculators, which are not interested in physical delivery, they tend to close their positions shortly before termination of the contract. They then enter into a new contract, whose date of delivery is further away in time. This period of liquidity transfer is commonly known as the rollover period. For further details of forwards and futures contracts, see Björk.[2]

2.2 The order book

The futures contracts from the preceding section are sold on an exchange. Such an exchange is represented by a limit order book, which contains the prices that market participants are willing to pay / receive for trading with the asset. If a buy limit order is placed, this represents the willingness of an investor to buy at the price associated with the limit order. If a seller enters a short position on the contract at the same price or lower, the trade will be executed. Hence, the futures price is at all times defined by the market - and the market information is summarized in the order book. An order book illustration can be seen in figure 2.1

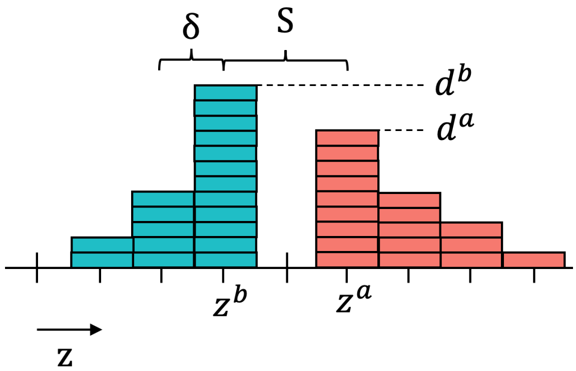


Figure 2.1: Illustration of an order book. The price axis runs horizontally and the number of orders that are queued for a specific price is shown on the y axis. The book has a tick size δ and spread S. The buy orders are queued on the left side - and the sell orders are queued on the right side. The buy order with the highest price z^b is called the bid price. The sell order with the lowest price z^a is called the ask price. The number of orders that are queued at z^a is called the ask depth d^a. The number of orders that are queued at z^b is called the bid depth d^b.

Note from figure 2.1 that the market has a smallest price increment δ, which is known as the "tick size". The buy limit orders in the market are represented by the blue boxes to the left and the sell limit orders are represented by the red boxes to the right. The highest price that any buyer is currently willing to pay for the asset is called the bid price - and it is denoted by z^b. Similarly, the lowest price of any sell limit order in the market is known as the ask price z^a. The number of limit orders that are waiting at the ask and bid prices are known as the first order depths of the market, denoted d^a and d^b respectively. The orders at the best ask and bid prices are said to be at the market. The buy limit orders that have a price lower than z^b and the sell limit orders that have a price higher than z^a are said to be out of the market. The data associated with orders that are outside the market is commonly referred to as higher order data.

The spread S, which in figure 2.1 is two, represents the number of ticks between the ask price and the bid prize. Any limit order that falls between the bid price and the ask price is said to be in the market. The arrival of such an order will change either the bid or the ask level, causing the new order to be at the market, and the previous orders on the same side to be out of the market.

Given an order book like the one in figure 2.1, a natural question arises. What is the current price of the asset? Moreover, given the full history of order books, what can be said about the assets fair value? Having perfect knowledge of the two allows an investor to capitalize on statistical arbitrage opportunities that arise in the market. However, neither of these quantities are measurable in a trivial way.

2.3 Market microstructure and the break down of Black Scholes

The fair price of any asset is determined by the expectations of its future price, discounted to account for the time-devaluation of money. This incentivises speculators to establish quantitative models that capture the future price moves of a given asset. The most popular way of quantifying prices of financial derivatives is through the Black Scholes framework. In this framework, prices are assumed to follow Geometric Brownian Motions (GBMs) with constant drift μ and volatility σ^2. This assumption makes the log-returns of the asset values normally distributed, leading to manageable calculations that can be performed analytically. In reality, the the log returns will not be normally distributed. However, if the time between measurements is sufficiently large, the returns will be perceived as log normal - which is a consequence of the central limit theorem. [3] This means that for small enough time scales, the Black Scholes framework will not result in an accurate model of the price increments.

Instead, the price can be thought of as an unobservable process that diffuses between the bid and the ask price. The value could theoretically venture out of this interval, but this would present an arbitrage opportunity. Such opportunities are generally capitalized on very quickly, making them highly unlikely. Every market event, defined as anything that changes the state of the order book, conveys information that will affect the fair price according to some distribution. By quantifying this distribution, one can gain insights into the concluding questions of section 2.2

2.4 Pricing a fill or kill limit order

Consider a fill or kill buy limit order being placed at time t. The order takes some time to reach the market, which it does at time T. The latency $T-t$ varies between market participants due to differences in bandwidth, physical distance from the exchange etc. However, it generally ranges from milliseconds to tenths of seconds.

If the limit order is executed with a limit price l and the fair value of the asset is z_T. Then the cost of placing that particular order would be $l - z_T$. This cost would be incurred at the future time T - and would normally be associated with a discount factor. However, since the time intervals of interest are in the magnitude of seconds, the overall drift of the asset price becomes irrelevant.[4]

If the order is killed, no payment is made on behalf of the trader. In such a scenario, the cost of the order is defined to be 0. Depending on the investor preferences, a penalty could also be imposed to represent the opportunity cost of not performing a trade. However, for the remainder of this thesis, the preferences of the investor will be chosen such that the opportunity cost of not performing a trade is 0. Under this assumption, the cost function can be expressed explicitly as

$$C := \mathbb{E}[l - z_T | e_T, \mathscr{F}_t] \cdot \mathrm{P}(e_T | \mathscr{F}_t) \tag{2.1}$$

Where $\mathscr{F}_t$ denotes the non-anticipating filtration, encompassing all available information about the asset at time t. Moreover, e_T symbolises the event that the order gets executed at time T. As described in section 1.5, the report only considers instances where the limit price is the same as the ask price at time t. In practice, we therefore have that $l = z_t^a$, where z_t^a indicates the ask price at time t. We also get that $e_T = I_{z_t^a \geq z_T^a}$, and (2.1) becomes

$$C = E[z_T^a - z_T | \{z_t^a \geq z_T^a\}, \mathscr{F}_t] \cdot \mathrm{P}(z_t^a \geq z_T^a | \mathscr{F}_t). \tag{2.2}$$

2.5 Estimates of the asset value

There are various ways to estimate the fair asset price from the time series of order books. One of the most trivial ways is through the mid price. This is simply the average between the ask and the bid level

$$M_t := \frac{z_t^a + z_t^b}{2} \tag{2.3}$$

Whilst simple, it suffers from a number of unsatisfactory properties. Firstly, the mid price only changes in the event that the ask price or bid price changes. It is therefore a low frequency signal, which becomes problematic if the bid and ask levels change infrequently. Secondly, it disregards all information prior to the time of evaluation. Thirdly, numerous empirical studies have shown that it is not a martingale. The reason is that consecutive ask price-movements tend to be negatively correlated.[5]

Assets can primarily be grouped into two classes, large-tick assets and small-tick assets. Large-tick assets are such that "the bid-ask spread rarely exceeds the minimum tick size".[6] These are the assets where the tick size is significant, and which therefore tend to trade with a spread equal to one. Another important characteristic of large-tick assets is that they tend to have large queues in the limit order book. The reason is that the price change of a one tick move will be highly significant. In contrast, for small-tick assets, the typical spread is much larger than one. In such case, investors can obtain priority in the order book by competing on price. This diminishes the importance of the position in the queue.[7]

A more sophisticated estimate of the fair price is the weighted mid price

$$W_t := z_t^b \cdot (1 - I_t) + z_t^a I_t \tag{2.4}$$

$$I_t := \frac{d_t^b}{d_t^a + d_t^b}, \tag{2.5}$$

where d_t^a is the ask depth and d_t^b is the bid depth at time t from figure 2.1. The imbalance I_t is a measure of the relative difference in the first-order depths d_t^a, d_t^b. This gives an indication of how the bid and ask levels are likely to evolve. A drawback of the weighted mid price is that it contains a large degree of noise - and similarly to the mid price, empirical studies have shown that it's not a martingale.[8] The weighted mid price can also decrease at the placement of an aggressive buy order, which is opposite to what one would expect.[8] Figure 2.2 shows a comparison between the mid price and the weighted mid price for a futures contract on the Australian Dollar.

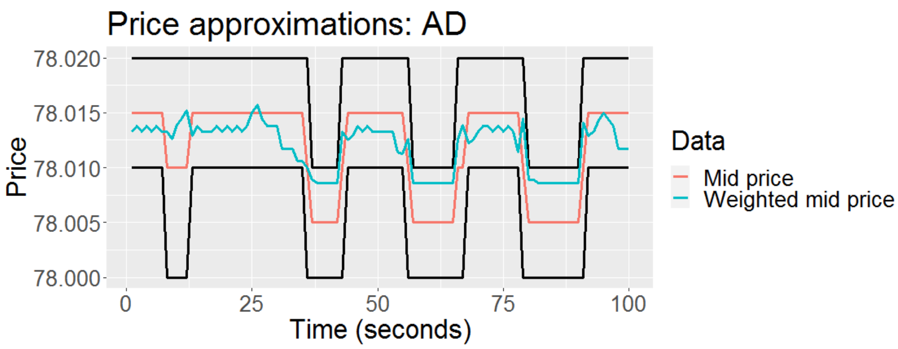


Figure 2.2: Illustration of the mid price and weighted mid price for Futures contracts on the Australian Dollar. The upper black line represents the ask level. The lower black line represents the bid level. The red line represents the mid price and the blue line represents the weighted mid price.

2.5.1 The Micro-price

The undesirable properties of the above estimators leaves one looking for better alternatives. One estimate, which is specifically constructed with the martingale property in mind, is Sotikov's micro-price.[8] The micro-price is defined as the mid price plus an adjustment term.

$$z_t^{micro} := M_t + g(I_t, S_t), \tag{2.6}$$

where I_t is the imbalance from (2.5) and $S_t = (z_t^a - z_t^b)/\delta$ is the spread. The micro-price can also be viewed as a sequence of approximations

$$z_t^{micro} = \lim_{i \to \infty} z_t^i. \tag{2.7}$$

The idea is for z_t^i to account for the first i mid price moves. To derive an expression for the micro-price, one relies on two assumptions. The first assumption is that the information in the order book is generated by the filtration from the following 3 dimensional Markov process

$$\mathscr{G}_t = \sigma(M_t, I_t, S_t). \tag{2.8}$$

The second assumption is that the expected mid-price increments are independent of the mid price level. This can be stated mathematically as

$$\begin{aligned}&\mathbb{E}_{\mathscr{G}_t}[M_{\tau_i} - M_{\tau_{i-1}} | M_t = M, I_t = I, S_t = S] \\ &\quad = \mathbb{E}_{\mathscr{G}_t}[M_{\tau_i} - M_{\tau_{i-1}} | I_t = I, S_t = S], \quad t \leq \tau_{i-1}.\end{aligned} \tag{2.9}$$

Under these assumptions, it can be shown that the ith mid-price prediction is

$$z_t^i = M_t + \sum_{k=1}^{i} g^k(I_t, S_t). \tag{2.10}$$

Where

$$g^1(I, S) = \mathbb{E}_{\mathscr{G}_t}[M_{\tau_1} - M_t | I_t = I, S_t = S] \tag{2.11}$$

and

$$g^{i+1}(I, S) = \mathbb{E}_{\mathscr{G}_t}\left[g^i\left(I_{\tau_1}, S_{\tau_1}\right) \mid I_t = I, S_t = S\right], \forall i \geq 0. \tag{2.12}$$

Note that $I \in [0, 1]$ and $S \in \mathbb{N}_+$. By discretizing these domains into n and m values respectively, one can simplify the expression of the micro price. Define the process $X_t := (I_t, S_t)$, along with the following matrices

$$\begin{cases} Q_{xy} & := & \mathrm{P}\left(M_{t+1} - M_t = 0 \wedge X_{t+1} = y \mid X_t = x\right) \\ R_{xk} & := & \mathrm{P}\left(M_{t+1} - M_t = k \mid X_t = x\right) \\ T_{xy} & := & \mathrm{P}\left(M_{t+1} - M_t \neq 0 \wedge X_{t+1} = y \mid X_t = x\right) \\ G^1(x) & := & \left(\sum Q^{s-1} R\right) K = (\mathbb{I} - Q)^{-1} R K \\ B & := & (\mathbb{I} - Q)^{-1} T \\ K & := & [-1, \ -\frac{1}{2}, \ \frac{1}{2}, \ 1]^T, \end{cases} \tag{2.13}$$

Where $Q \in R^{nm \times nm}$, $R \in R^{nm \times 4}$, $T \in R^{nm \times nm}$, $B \in R^{nm \times nm}$, $\mathbb{I}$ is the $nm \times nm$ identity matrix and $k \in K$. G^1 is a vector of length nm, where each element represents the first order adjustment for a particular imbalance and spread. The matrix B is used to express the ith mid price prediction, which becomes

$$z_t^i = M_t + (\sum_{k=0}^{i} B^k G^1) \cdot \vec{e}_{I_t, S_t}. \tag{2.14}$$

The vector $\vec{e}_{I_t,S_t}$ is the indicator vector, whose non-zero element corresponds to the active state of I_t and S_t. The micro-price is attained by letting the number of adjustments tend towards infinity.

$$z_t^{micro} = \lim_{i\to\infty} z_t^i = \lim_{i\to\infty} M_t + (\sum_{k=0}^{i} B^k G^1) \cdot \vec{e}_{I_t,S_t}. \tag{2.15}$$

A caveat of the above expression is that the sum in (2.15) is not guaranteed to converge. But as the following theorem states, it will if certain conditions are met.

Theorem 1 *If $B^* = \lim_{k\to\infty} B^k$ satisfies $B^*G^1 = 0$, then the limit*

$$\lim_{i\to\infty} z_t^i = z_t^{micro} \tag{2.16}$$

converges.

In practice, the condition posed by theorem 1 can be achieved by a pre-processing of the data. This involves symmeterizing the data so that every observed state (characterized by $(I_t, S_t, I_{t+1}, S_{t+1}, dM)$, where $dM := M_{t_i} - M_{t_{i-1}}$) is complemented by a mirrored state $(1 - I_t, S_t, 1 - I_{t+1}, S_{t+1}, -dM)$. This ensures that $B^*G^1 = 0$, which leads to the converging poperty as per theorem 1. For further details, as well as a proof of theorem 1, please refer to Stoikov.[8]

2.6 Modeling order book evolution

The expression for the limit order cost (2.2) can be further expanded by conditioning on whether the ask level changes in the interval $[t, T]$.

$$\begin{aligned} C = &\mathbb{E}[z_T^a - z_T | \{z_t^a = z_T^a\}, \mathscr{F}_t] \cdot P(\{z_t^a = z_T^a\} | \mathscr{F}_t) \\ &+ \mathbb{E}[z_T^a - z_T | \{z_t^a > z_T^a\}, \mathscr{F}_t] \cdot P(\{z_t^a > z_T^a\} | \mathscr{F}_t) \end{aligned} \tag{2.17}$$

Note that there is no term for $\{z_t^a < z_T^a\}$ since an increase in ask level would mean no execution. Establishing the cost of a fill or kill limit order can therefore be broken down into determining the above four terms. The first product in (2.17) will dominate the expression, especially for small times. This is because the transition probability $P(\{z_t^a > z_T^a\} | \mathscr{F}_t) \to 0$ as $[T - t] \to 0$.

The non-anticipating filtration $\mathscr{F}_t$ is assumed to be well approximated by the filtration $\mathcal{F}_t^{\mathscr{O}}$. $\mathcal{F}_t^{\mathscr{O}}$ is generated by the sequence of historical order books $\vec{\mathscr{O}}_t = \{\mathscr{O}_1, \mathscr{O}_2, \ldots \mathscr{O}_k\}$, where $\mathscr{O}_k$ is the order book which is being observed at time t. We therefore have

$$\mathcal{F}_t^{\mathscr{O}} := \sigma \{\mathscr{O}_1, \mathscr{O}_2, \ldots \mathscr{O}_k\}. \tag{2.18}$$

Also note that $\mathscr{G}_t \in \mathcal{F}_t^{\mathscr{O}}$. The use of this filtration leads to a slightly modified version of (2.17), which takes on the form

$$\begin{aligned} C = &\mathbb{E}[z_T^a - z_T | \{z_t^a = z_T^a\}, \mathcal{F}_t^{\mathscr{O}}] \cdot \mathrm{P}(\{z_t^a = z_T^a\} | \mathcal{F}_t^{\mathscr{O}}) \\ &+ \mathbb{E}[z_T^a - z_T | \{z_t^a > z_T^a\}, \mathcal{F}_t^{\mathscr{O}}] \cdot \mathrm{P}(\{z_t^a > z_T^a\} | \mathcal{F}_t^{\mathscr{O}}). \end{aligned} \tag{2.19}$$

In light of equation 2.19, one becomes interested in modeling the short term behaviour of the limit order book. This behaviour is assumed to be separable into one part which models the market activity, i.e. the incoming market event frequency. The other models the transitional behaviour of the order book, conditioned on a certain number of market events. The limit order book will be assumed to follow a Markov process, whose transition probabilities are only affected by the previous order books - and not the time in between market events. The theory involving these models will be discussed in the preceding chapters.

2.7 Markov chains

A Markov chain is a stochastic process $\{X_t\}$, describing a sequence of possible events, where the probability of moving to a certain state only depends on the preceding state. In other words

$$\mathrm{P}(X_t = x | X_1 = x_1, X_2 = x_2, , X_{t-1} = x_{t-1}) = \mathrm{P}(X_t = x | X_{t-1} = x_{t-1}) \tag{2.20}$$

A Markov chain is said have finite memory if the number of attainable states are finite. The one step transition probabilities of a discrete Markov chain with finite state space are given by the transition matrix Π, which satisfies

$$\Pi_{ij} := \mathrm{P}\left(X_{n+1} = j \mid X_n = i\right) \tag{2.21}$$

Let $\mu_n = [p_1, p_2, \ldots, p_m]$, where $\sum_{i=1}^{m} p_i = 1$, be some probability distribution of markov states in a finite state space. Then the distribution μ_{n+k}, which occurs k steps later, can be expressed as

$$\mu_{n+k} = \mu_n \Pi^k. \tag{2.22}$$

An absorbing state in a Markov chain is a state that, once entered, cannot be left. In other words, if i is an absorbing state, then $\Pi_{i,i} = 1$. Therefore, if the state representation is ordered such that the r absorbing states come last, one gets a transition probability matrix of the form.

$$\Pi = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix} \tag{2.23}$$

Where Q is a rectangular transition matrix between transient states, R is a transition matrix from transient to absorbing states and I is the identity matrix.

2.8 Modeling the order book by a Markov chain

The order book is modeled by a Markov Chain. In order to ensure finite memory and stationairty, changes in parameters values, as opposed to the parameters themselves, are considered. To this end, the following definition is made

$$\zeta(x_t, x_{t-1}) := \begin{cases} 1 & \text{if } \ x_t > x_{t-1} \\ 0 & \text{if } \ x_t = x_{t-1} \\ -1 & \text{if } \ x_t < x_{t-1}. \end{cases} \tag{2.24}$$

The state of the Markov process is represented as

$$S(t_i) := \begin{bmatrix} I_{t_i} \\ \zeta(d^a_{t_i}, d^a_{t_{i-1}}) \\ \zeta(d^b_{t_i}, d^b_{t_{i-1}}), \end{bmatrix}^{\mathrm{T}}, \tag{2.25}$$

where I_{t_i} is the imbalance at time t_i. In order to reduce the state-space, I_{t_i} was binned according to $\{[0, 0.25), [0.25, 0.5), [0.5, 0.75), [0.75, 1]\}$. Moreover, two absorbing states were defined to represent when the ask price changes. One for the case when the ask level increases - and a separate one for when the ask level decreases. The state space therefore contained a total of $4*3*3+2 = 38$ states. With help of the following hash function, the state was represented as an identity vector

$$\mu_{t_i} = \begin{cases} e(37) & \text{if } \ \zeta(z^a_{t_i}, z^a_{t_{i-1}}) = 1 \\ e(5 + \zeta(d^b_{t_i}, d^b_{t_{i-1}}) + 3\zeta(d^a_{t_i}, d^a_{t_{i-1}}) + 9\lfloor \frac{I_{t_i}}{0.25} \rfloor) & \text{if } \ \zeta(z^a_{t_i}, z^a_{t_{i-1}}) = 0 \\ e(38) & \text{if } \ \zeta(z^a_{t_i}, z^a_{t_{i-1}}) = -1, \end{cases} \tag{2.26}$$

where $e(x), \ x \in 38$ is the indicator vector, with the xth element equal to one. This gives a transition probability matrix π of size 38×38 of the following form

$$\Pi = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix}, \tag{2.27}$$

where $Q \in \mathbb{R}^{36\times36}$, $R \in \mathbb{R}^{36\times2}$ and $I \in \mathbb{R}^{2\times2}$.

2.9 Bayesian inference of Markov chains

Bayesian inference is used to derive predictions of model parameters with the help of historical observations. Let ξ be a parameterized model, which has been postulated to capture the behaviour of the data. Let $\vec{\theta}$ represent the parameters of this model and let $\vec{x}$ be a vector of observations. Bayes theorem states that

$$\mathrm{P}(\vec{\theta}|\vec{x}, \xi) = \frac{\mathrm{P}(\vec{x}|\vec{\theta}, \xi)\,\mathrm{P}(\vec{\theta}|\xi)}{\mathrm{P}(\vec{x}|\xi)} \tag{2.28}$$

The so called model evidence $\mathrm{P}(\vec{x}|\xi)$ is the same for all possible parameter values. It therefore does not affect the relative probabilities of the different model parameters. Instead, it serves as a normalization term and makes sure that $\mathrm{P}(\vec{\theta}|\vec{x}, \xi)$ satisfies the criteria of a distribution. This term tends to be very hard to evaluate, which in turn hinders the evaluation of the posterior distribution $\mathrm{P}(\vec{\theta}|\vec{x}, \xi)$. This issue can be amended is by making use of conjugate priors.

Conjugate priors cause the posterior distribution $\mathrm{P}(\vec{\theta}|\vec{x},\xi)$ to take on the same form as the prior distribution $\mathrm{P}(\vec{\theta}|\xi)$, thereby removing the need to explicitly calculate the model evidence. For the Markov chain, described by the transition probability matrix in section 2.8, the conjugate prior is a product on Dirichlet distributions.[9] Moreover, the parameter vector $\vec{\theta}$ is

$$\vec{\theta} := \{\alpha(\mu_{t_i}|\mu_{t_{i-1}}),\ \ i \in [1,2,\ldots,N]\}, \tag{2.29}$$

where each hyperparameter $\alpha(\mu_{t_i}|\mu_{t_{i-1}}) \in \mathbb{R}_+$. The quantity $\mathrm{P}(\vec{\theta}|\xi)$ is known as the prior distribution - and it can be used to account for previous assumptions on the model dynamics. Moreover, the initial values that one assigns so the hyperparameters can be thought of as a set of fake counts $\tilde{n}(\mu_{t_i},\mu_{t_{i-1}})$, such that $\alpha(\mu_{t_i},\mu_{t_{i-1}}) = \tilde{n}(\mu_{t_i},\mu_{t_{i-1}}) + 1$. By setting $\{\alpha(\mu_{t_i}|\mu_{t_{i-1}}) = 1,\ \forall \mu_{t_i} \in U\}$, which is equivalent to setting all fake counts to be 0, one gets a uniform prior distribution over the model parameters. This represents the scenario where no prior information about the distribution is taken into account.

With the use of such a conjugate prior, the posterior distribution for the Markov chain parameters can be analytically expressed as

$$\begin{aligned}\mathrm{P}(\theta|\vec{x},\xi) = \prod_{\mu_{t_i}\in U} \Bigg\{ & \frac{\Gamma(n(\mu_{t_{i-1}}) + \alpha(\mu_{t_{i-1}}))}{\prod_{\mu_{t_{i-1}}\in U} \Gamma(n(\mu_{t_i},\mu_{t_{i-1}}) + \alpha(\mu_{t_i},\mu_{t_{i-1}}))} \\ & \times \delta(1 - \sum_{\mu_{t_{i-1}}} p(\mu_{t_i}|\mu_{t_{i-1}})) \\ & \times \prod_{\mu_{t_{i-1}}\in U} p(\mu_{t_i}|\mu_{t_{i-1}})^{n(\mu_{t_i},\mu_{t_{i-1}})+\alpha(\mu_{t_i},\mu_{t_{i-1}})-1} \Bigg\},\end{aligned} \tag{2.30}$$

where $n(\mu_{t_i},\mu_{t_{i-1}})$ is the number of observations going from $\mu_{t_{i-1}}$ to μ_{t_i}. Moreover, equation (2.30) makes use of the shorthand notations $n(\mu_{t_i}) := \sum_{(\mu_{t_{i-1}}\in U)} n(\mu_{t_i},\mu_{t_{i-1}})$ and $\alpha(\mu_{t_i}) := \sum_{(\mu_{t_{i-1}}\in U)} \alpha(\mu_{t_i},\mu_{t_{i-1}})$. Since the posterior also follows a product of Dirichlet distributions, the mean can be analytically expressed as

$$\mathbb{E}^{\mathrm{post}}_{\mathcal{F}^{\mathcal{O}}_{t_i}}[\mathrm{P}(\mu_{t_i}|\mu_{t_{i-1}})] = \frac{n(\mu_{t_i},\mu_{t_{i-1}}) + \alpha(\mu_{t_i},\mu_{t_{i-1}})}{n(\mu_{t_{i-1}}) + \alpha(\mu_{t_{i-1}})}. \tag{2.31}$$

Equation 2.31 can be factored into

$$\begin{aligned}\mathbb{E}^{\mathrm{post}}_{\mathcal{F}^{\mathcal{O}}_{t_i}}[\mathrm{P}(\mu_{t_i}|\mu_{t_{i-1}})] = & \frac{1}{n(\mu_{t_{i-1}}) + \alpha(\mu_{t_{i-1}})} \\ & \times \left[n(\mu_{t_{i-1}}) \left(\frac{n(\mu_{t_{i-1}},\mu_{t_{i-1}})}{n(\mu_{t_{i-1}})} \right) + \alpha(\mu_{t_{i-1}}) \left(\frac{\alpha(\mu_{t_i},\mu_{t_{i-1}})}{\alpha(\mu_{t_{i-1}})} \right) \right]\end{aligned} \tag{2.32}$$

where $n(\mu_{t_i},\mu_{t_{i-1}})/n(\mu_{t_{i-1}})$ is the maximum likelihood estimator of the given data and $\alpha(\mu_{t_i},\mu_{t_{i-1}})/\alpha(\mu_{t_{i-1}})$ is the prior expectation of transitions going from $\mu_{t_{i-1}}$ to

μ_{t_i}. Note that if $n(\mu_{t_{i-1}}) >> \alpha(\mu_{t_{i-1}})$, the posterior estimate will converge to the MLE. Lastly, the variance of the posterior follows

$$\mathrm{Var}^{\mathrm{post}}_{\mathcal{F}^{\mathcal{O}}_{t_i}}(\mathrm{P}[(\mu_{t_i}|\mu_{t_{i-1}})]) = \frac{n(\mu_{t_i},\mu_{t_{i-1}}) + \alpha(\mu_{t_i},\mu_{t_{i-1}})}{(n(\mu_{t_{i-1}}) + \alpha(\mu_{t_{i-1}}))^2} \times \frac{(n(\mu_{t_{i-1}}) + \alpha(\mu_{t_{i-1}})) - (n(\mu_{t_i},\mu_{t_{i-1}}) + \alpha(\mu_{t_i},\mu_{t_{i-1}}))}{(n(\mu_{t_{i-1}}) + \alpha(\mu_{t_{i-1}}) + 1)}. \tag{2.33}$$

2.10 Modeling market activity

Since the number of incoming market events depends on the number of active market participants, one would expect the market activity to fluctuate throughout the day. Moreover, once an order is placed, another one is more likely to follow soon thereafter. This is based on the assumption that other market participants are more likely to readjust their positions shortly after a trade has entered the market.

The simplest model of market activity would be a constant frequency f_0. The model can be made more sophisticated by taking the time of day, which is represented by $\nu \in [0, 24)$, into account. The market activity can therefore be modeled as.

$$f_1(\nu) := f_0 \cdot g(\nu) \tag{2.34}$$

Where the function $g(\nu)$ satisfies $\mathbb{E}[g(\nu)] = 1$. An additional variable of importance is the time Δt since the last order. With this in mind, the frequency model $f_2(\nu, \Delta t)$ is defined. The function is assumed to be separable in its parameters - and the market activity is therefore modeled as

$$f_2(\nu, \Delta t) := f_0 \cdot g(\nu) \cdot h(\Delta t), \tag{2.35}$$

where $\nu \in [0, 24)$, $\Delta t \in [0, \infty)$ and the functions $g(\nu)$ and $h(\Delta t)$ satisfy $\mathbb{E}_{\mathcal{F}^{\mathcal{O}}_{t_i}}[g(\nu)] = \mathbb{E}_{\mathcal{F}^{\mathcal{O}}_{t_i}}[h(\Delta t)] = 1$. However, this instantaneous model of the market activity does not give full insight into the average market activity that an investor observes during the time interval $[t, T]$. To take the latency period into account, a final model is defined.

$$f_3(\nu, \Delta t, T - t) := f_0 \cdot g(\nu) \cdot H(\nu, \Delta t, T - t), \tag{2.36}$$

where $H(\nu, \Delta t, T - t)$ is a modification of $h(\Delta t)$ which takes the latency period into account. It also satisfies $\mathbb{E}_{\mathcal{F}^{\mathcal{O}}_{t_i}}[H(\nu, \Delta t, T - t)] = 1$.

2.10.1 Activity fluctuations thoughout the day

The assets of interest are traded on CME Group. They are traded around the clock, apart from a one hour daily break between [16:00, 17:00] central time (CT). However, the trading activity tends to be substantially lower during the intervals 15:00 -

16:00 and 17:00 - 18:00. This is a consequence of the time zone differences between countries varying throughout the year. To see why, consider a Japanese investor starting his day at 08:00 (JST). This would either correspond to 17:00 or 18:00 in (CT). Which one it is depends on the daylight savings time cycle in the US. The consequence is that the trader is always active at $\nu = 18$, but only sometimes for $\nu = 17$.

Apart from the frequency drop discussed above, market activity displays three significant peaks throughout the day, which correspond to a high trading activity in America, Europe and Asia. The American peak tends to be the most prominent, followed by the European and lastly the Asian. To capture these behaviours, the function $g(\nu)$ was parameterized according to

$$\begin{cases} g(x; \vec{a}, \vec{\mu}, \vec{\gamma}, c) := R(x) \cdot (a_1 f_{\mu_1,\gamma_1}(x) + a_2 f_{\mu_2,\gamma_2}(x) + a_3 f_{\mu_3,\gamma_3}(x) + k) \\ R(x) := I\{x \in [0, 24) \cap [16, 17]\} \cdot (1 - c \cdot I\{x \in [15, 18]\}) \\ \text{such that} \\ \mathbb{E}[g(u)] = 1, \text{ where } u \sim U[0, 24), \end{cases} \tag{2.37}$$

where $f_{x_0,\gamma}(x)$ is the Cauchy probability density function with location parameter x_0 and shape parameter γ, i.e.

$$f_{x_0,\gamma}(x) = \frac{1}{\pi\gamma[1 + (\frac{x - x_0}{\gamma})^2]}. \tag{2.38}$$

An analytical expression for the constant k is derived in A.1.

2.10.2 Time since last market event

Market makers and other market participants change their positions based on new information in the market, which comes in the form of orders placed by other investors. For this reason, one would expect the market activity to be higher shortly after an order has been placed. Similarly, one would expect a low market activity to persist if the time since the last market event is large. To model this behaviour, a function $h(\Delta t)$ is introduced. It represents the frequency compensation one would expect, up until the next market event, given that the last event occured a time Δt ago. This decaying function is parameterized as

$$h(\Delta t) := \frac{a}{b + \Delta t} + c. \tag{2.39}$$

However, if a new market event arrives within the latency period $[t, T]$, the function $h(\Delta t)$ will no longer model the average market event experienced by the investor. To account for this, the function $h(\Delta t)$ is replaced by $H(\nu, \Delta t, T - t)$. Under the simplifying assumption that the frequency is deterministically modeled by the function $h(\Delta t)$, one can derive an expression for $H(\nu, \Delta t, T - t)$. The detains of this derivation are presented in A.2, and its resulting parameterization is

$$H(\nu, \Delta t, T - t) = \begin{cases} h(\Delta t) \ , \quad \text{if} \quad h(\Delta t) \leq x \\ h(0) + x(1 - \frac{h(0)}{h(\Delta t)}) \ , \quad \text{if} \quad h(\Delta t) > x \\ x = 1/(T - t) g(\nu) \ \mathrm{f}_0. \end{cases} \tag{2.40}$$

2.10.3 Putting it together

Assume that the number of market events in the interval $[t, T]$ is some whole number n. The transition probability term $\mathrm{P}(\{z^a_t > z^a_T\}|\mathcal{F}^{\mathscr{O}}_t)$ in equation (2.19) could then be approximated as

$$\mathrm{P}(\{z^a_t > z^a_T\}|\mathcal{F}^{\mathscr{O}}_t) \approx \cdot\mu \cdot \Pi^n[.,j]. \tag{2.41}$$

Where j is such that $\Pi^n[.,j]$ corresponds to a n-step transition to the absorbing state $\{z^a_{i+1} < z^a_i\}$. The number of market events n was approximated by the frequency distribution derived in section 2.10. Note that this expectation need not be a whole number. Therefore, the final approximation of the transition probability is interpolated according to

$$\begin{aligned} \mathrm{P}(\{z^a_t > z^a_T\}|\mathscr{F}_t) \approx \mu \cdot \Big((\lceil\eta\rceil - \eta)\Pi^{\lfloor\eta\rfloor}[.,j] + (\eta - \lfloor\eta\rfloor)\Pi^{\lceil\eta\rceil}[.,j]\Big)\,. \\ \eta = \mathbb{E}_{\mathcal{F}^{\mathscr{O}}_{t_i}}[f_i(x)]. \end{aligned} \tag{2.42}$$

Where $f_i(x)$ is any of the frequency distribution models from section 2.10, $\lfloor x\rfloor$ is the floor function and $\lceil x\rceil$ is the ceiling function.

3

Methods and Results

This chapter covers the evaluation procedure of the models described in chapter 2. Unless otherwise stated, the functions in this chapter have been fitted by means of the least squares method - and the empirically fitted version of any model m will be referred to as $\hat{m}$. Whenever the testing and training data is mentioned, it is referring to the data partitioning described in section 3.1. A number of figures will display model MSEs. These have been made relative, meaning that they show the relative MSE with respect to some benchmark model. This benchmark model is stated for each graph individually. The corresponding absolute MSE plots can be found in section A.4.

The following sections will make frequent use of references to the last market event before some time t. It will also refer to the first market event preceding some time t. The times of these events are denoted $\underline{t}$ and $\bar{t}$ respectively. They are formally defined as

$$\begin{cases} \underline{t} & := \max \tilde{t}_i \quad \text{s.t.} \quad \tilde{t}_i \in \vec{\tilde{t}}, \quad \tilde{t}_i \leq t \\ \bar{t} & := \min \tilde{t}_i \quad \text{s.t.} \quad \tilde{t}_i \in \vec{\tilde{t}}, \quad \tilde{t}_i > t. \end{cases} \tag{3.1}$$

An illustration of market event times, along with $\underline{t}$ and $\bar{t}$, can be seen in figure 3.1.

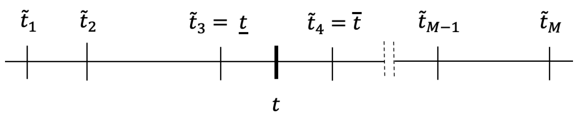


Figure 3.1: Illustration of market event times. There are M market events, occurring at times $\{\tilde{t}_1, \ldots, \tilde{t}_M\}$. The time t is arbitrarily selected and does not necessarily correspond to a market event. $\underline{t}$ is the time of the last market event before t. $\bar{t}$ is the time of the first market event preceding t.

3.1 Data

The data used in the analysis came from Futures contracts on Australian Dollars, Canadian dollars, Brittish pounds, and Japanese yen. These were all traded on CME Group between Sun 5:00PM - Fri 4:00 PM Central Time (CT) - and there was a daily one hour break between 16:00 and 17:00 (CT). Since the market opening hours are expressed in CT, it will be the default time zone for all times mentioned.

As described in section 2.1, futures contracts have a delivery date. On CME group, any currency is associated with four annual contracts. This leads to four roll over periods, where liquidity moves from one contract to the next. The roll overs happen in a relatively fast and synchronized manner. Unreliable results as a consequence of this rollover period were avoided by removing data from the day leading up to and preceding the rollover.

The dataset ranged from 2015-03-01 to 2018-10-17. During this time, market data was registered with nanosecond precision. The dataset was split into one training set and one testing set. The training set ranged from 2015-03-01 to 2017-12-31, whereas the testing set ranged from 2018-01-01 to 2018-10-17. All the models were fitted using the training dataset and all testing was conducted on the testing dataset, which ensured that causality was not violated.

3.2 Market activity models

The base-frequency f_0 of the contract was estimated as the mean frequency $\hat{f}_0$ of the training set. $\hat{f}_0$ was evaluated by dividing the total number of market events by the duration of the training data set (the daily one hour break was considered part of the active interval).

Figure 3.2 shows $\hat{g}(\nu)$, the fitted version of $g(\nu)$ from section 2.10.1. $\hat{g}(\nu)$ was fitted using MLE on a histogram of data with 1440 intervals. The figure also contains the training and testing data. Note that $\hat{g}(\nu)$ largely captures the behaviour of the testing data. Moreover, the market activity goes to 0 between 16:00 - 17:00. The trading activity is also substantially lower during the intervals 15:00 - 16:00 and 17:00 - 18:00, Which is well captured by the parameterization. Lastly, the existence of outliers can be noted around 10:00, which are associated with a significant increase in market activity.

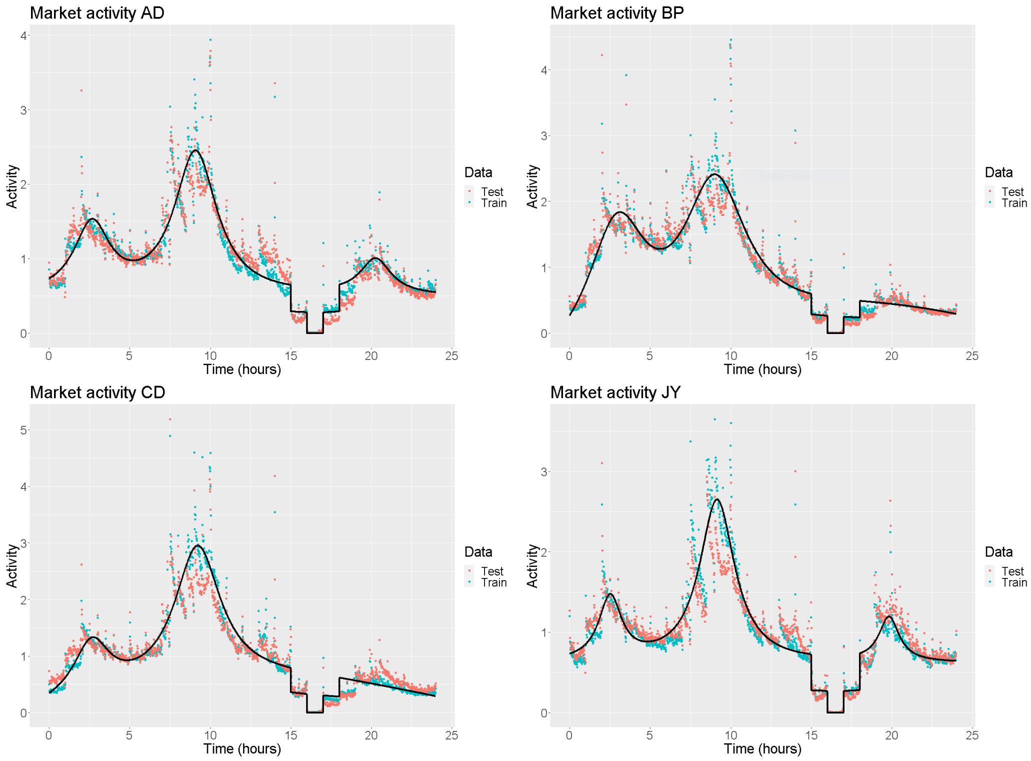


Figure 3.2: Market activity as a function of the time of day ν. The figure contains training data (blue), testing data (red) and $\hat{g}(\nu)$ - the fitted version of $g(\nu)$ from section 2.10.1.

The data collection procedure for $\hat{h}(\Delta t)$ is illustrated by figure 3.3. Time points $\{t_1, \ldots t_N\}$ were selected with a fixed spacing $I = 100\,\text{s}$. For each time point t_n, the times $\{\underline{t_n}, \overline{t_n}\}$ of the previous and receding events were recorded. Δt was measured as $t_n - \underline{t_n}$.

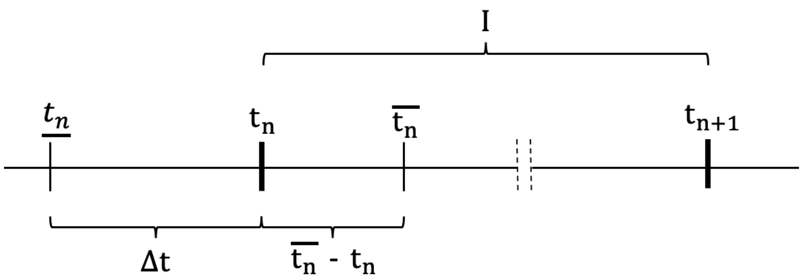


Figure 3.3: Data collection procedure for $h(\Delta t)$, with time ranging over the horizontal axis. The times t_n, $n \in \{1, \ldots, N-1\}$ were predetermined.

From here, the estimate $\hat{h}(\Delta t)$ was calculated as $\hat{h}(\Delta t) = 1/[(\overline{t_n} - t_n) \cdot \hat{g}(\nu) \cdot \hat{f}_0]$, where $\hat{f}_0$ and $\hat{g}(\nu)$ were evaluated according to the preceding discussion. Figure 3.4 shows the fitted function $h(\Delta t)$, along with the training and testing data.

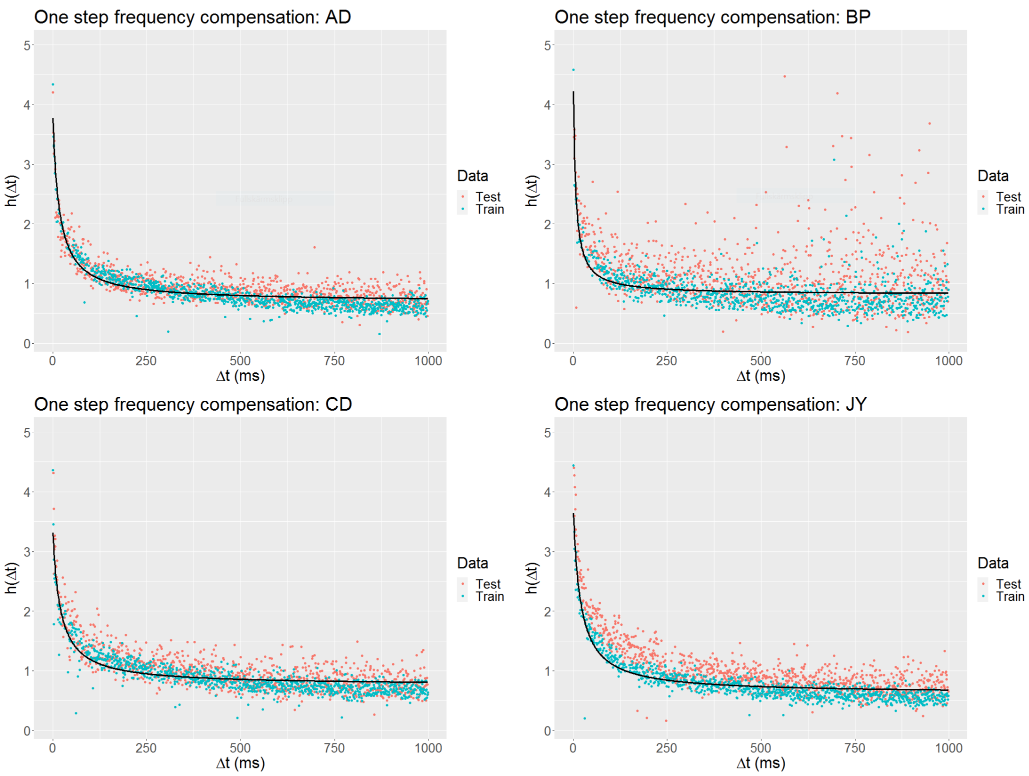


Figure 3.4: One step frequency compensation $h(\Delta t)$, as a function of the time Δt since the last order. The blue data comes from the training set - and the red data comes from the testing set.

Lastly, $\hat{H}(T-t, \Delta t)$ was determined from the parameterization of $\hat{h}(\Delta t)$, by means of equation (2.40).

3.2.1 MSE for the number of market events

Figure 3.5 illustrates the evaluation procedure used to compare the market activity models.

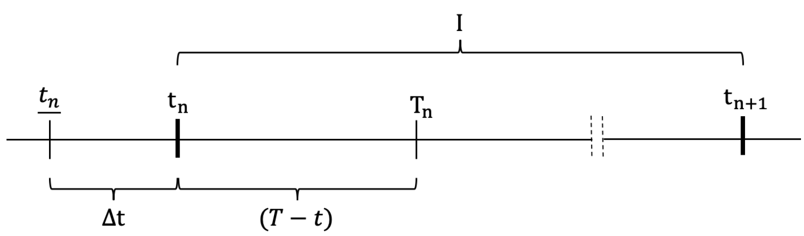


Figure 3.5: Evaluation procedure for market activity models. The time points $\{t_1, \ldots t_{N-1}\}$ were separated by a fixed interval $I = 3\,\text{min}$. The time $(T-t)$ was kept fixed throughout all intervals.

The testing data was divided into intervals of equal length $I = 3\,\text{min}$. For each interval, the time of the last preceding order $\underline{t_n}$ was recorded - and Δt was evaluated as $t_n - \underline{t_n}$. With this information, the fitted frequency models $\{\hat{f}_0, \ldots \hat{f}_3\}$ from section 2.10 could be evaluated. The time T_n was selected so that

$$T_n - t_n = (T - t), \quad \forall n \in \{1, \ldots, N\}, \tag{3.2}$$

where $(T - t)$ represents the desired latency period. The number of predicted orders for frequency model f_i was therefore $f_i \cdot (T - t)$. The corresponding residual was calculated as the difference between this prediction and the observed number of orders in the interval $[t_n, T_n]$. Figure 3.6 contains relative MSEs for the number of orders observed in an interval of 100 ms. The fitted fixed frequency model f_0 was used as the benchmark.

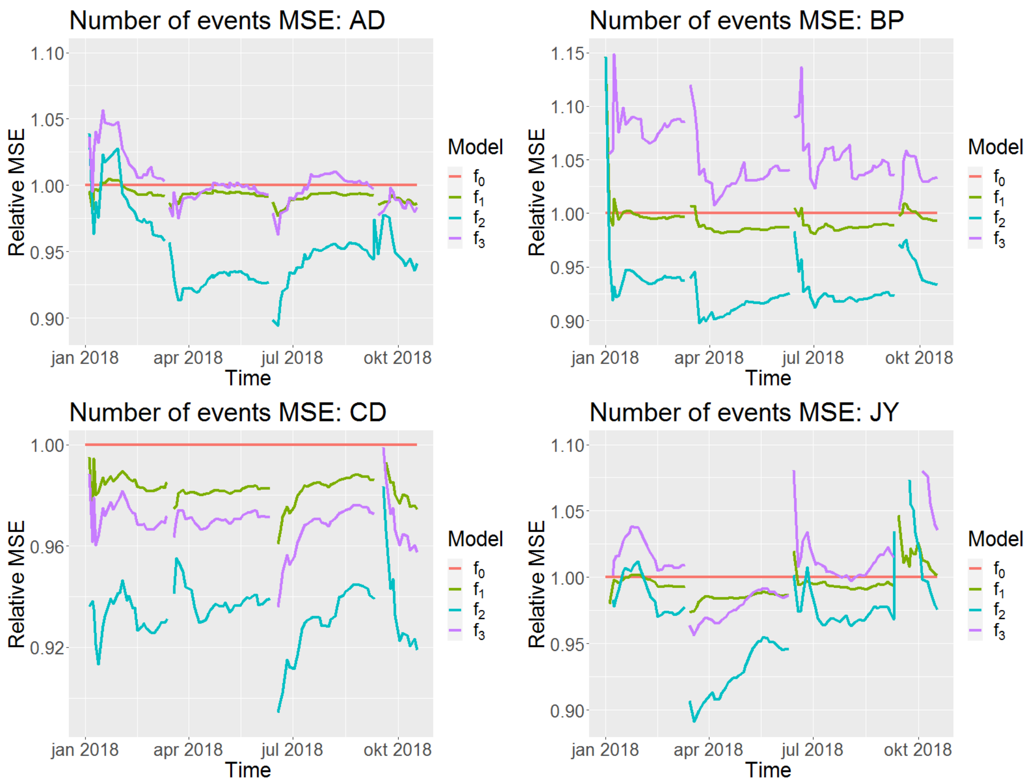


Figure 3.6: Relative MSEs for the number of orders observed in an interval of 100 ms. The MSE values are reset at the change of each contract, resulting in the three discontinuities across the time period.

Note that $\hat{f}_2$ consistently maintains the lowest MSE of the frequency models (except for short intervals of of the JY contracts). Note also that $\hat{f}_3$ tends to give results comparable to $\hat{f}_1$ and $\hat{f}_0$.

3.3 Markov transition models

As described in section 2.8, the order book was modeled by a Markov chain. This section describes the parameter fitting procedure, as well as the model validation procedure, associated with this assumption.

To quantify the degree by which the Markov transition model predicted changes in future ask levels, the full transition matrix was compared to two other models. A naive one, which contained constant up and down transition probabilities for all states, as well as a no-memory model. The no-memory model only took the imbalance into consideration and thus neglected previous market orders. All parameters of these models were extracted from the empirically estimated transition matrix $\hat{\Pi}$.

In order to fit the Markov transition matrix, all of the market events $\{\tilde{t}_2, \tilde{t}_3, \ldots, \tilde{t}_{M-1}\}$ contained in the training set were used. For each of these time points $\tilde{t}_n$, the market event at $\tilde{t}_{n-1}$ was used to evaluate the Markov state $\mu_{\tilde{t}_{i-1}}$. Moreover, the observations $\{\tilde{t}_n, \tilde{t}_{n+1}\}$ were used to evaluate $\mu_{\tilde{t}_i}$. With the help of these quantities, equation (2.31) was used to evaluate an estimate of the transition matrix. It should be noted that the instances for which $z^a_{\tilde{t}_n} \neq z^a_{\tilde{t}_{n-1}}$ were excluded from this evaluation. The reason being the assumption that the Markov state associated with a change in ask level is absorbing.

Once the full transition matrix had been evaluated, the two less detailed transition matrices could be inferred. The naive one only used a fixed transition probability for all states - and was therefore calculated as two weighted averages. One for the up transitions and one for the down transitions. These averages weighted the transition probability of each element in the transition matrix by the number of observations that were used to generate the value. An identical, but more fine grained approach was used to evaluate the no history model. Here, the entries in the transition matrix were grouped by imbalance - and the up and down transition probabilities were calculated for each of the imbalance values.

3.3.1 k-step ask level transition probabilities

The sampling procedure for evaluating the k-step ask level transition probabilities is illustrated in figure 3.7. Sample times t_n, $n \in \{1, \ldots, N-1\}$ were selected with a uniform spacing $I = 10\,\mathrm{s}$. For each of these times, the previous order books (at time $\tilde{t}_m = \underline{t_n}$ and $\tilde{t}_{m-1}$) were used to evaluate the Markov state. The three transition models were used to predict the probability that the ask level at time $\tilde{t}_{m+k}$ would be different from the one at time $\underline{t_n}$.

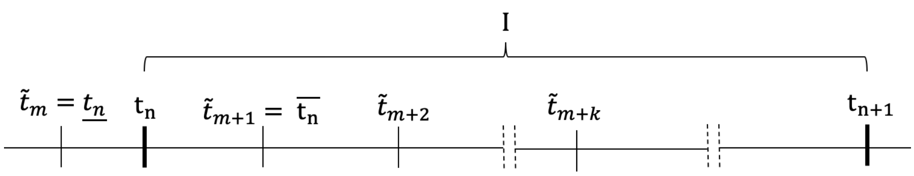


Figure 3.7: Evaluation procedure for the k-step transition probability of the Markov transition models.

The order book at time $\tilde{t}_{m+k}$ was then used to measure $I\{z^a_{t_2} < z^a_{t_1}\}$ and $I\{z^a_{t_2} > z^a_{t_1}\}$. Residuals were attained by subtracting the predicted transition probabilities from their respective indicator functions. The residuals were then used to calculate MSEs for the transition probabilities of the respective models. Figure 3.8 shows MSE plots for the one step ask level transition probabilities, i.e. $k = 1$. Note that the full and no history models tend to have a lower MSE than the naive model. Also note that there is no direct indication that the full model outperforms the no history model.

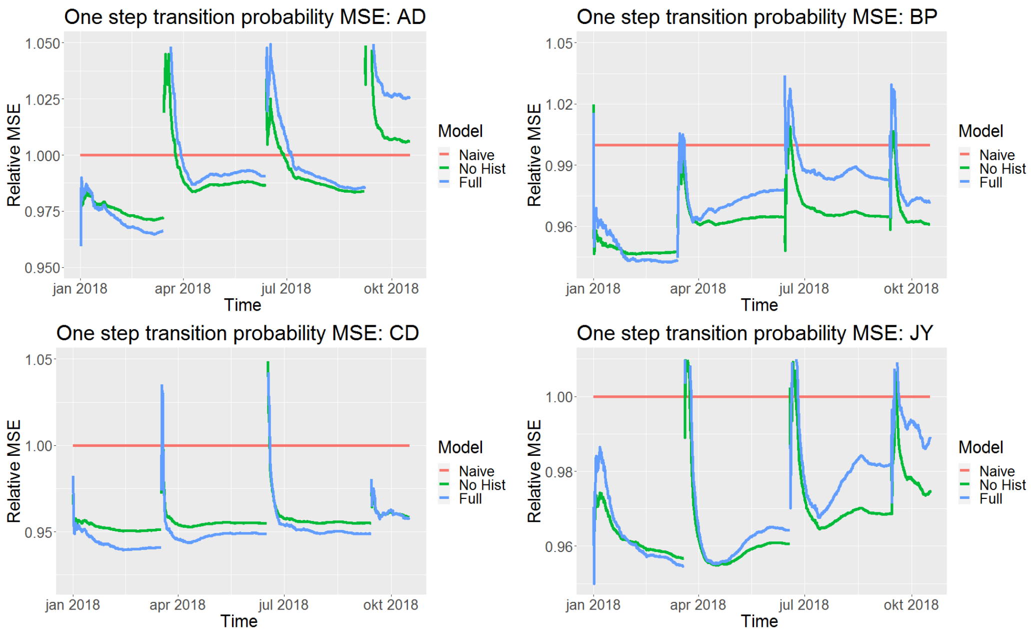


Figure 3.8: Relative MSE for the single transition probabilities of the assets. The MSE values are reset at the change of each contract, resulting in the three discontinuities across the time period.

Figure 3.9 shows relative MSE plots for the multiple step ask level transition probabilities ($k > 1$), with the naive model acting as a benchmark. Note that the no history model tends to continuously outperform the full model. Note also that both of these models outperform the naive model, with the exception of AD.

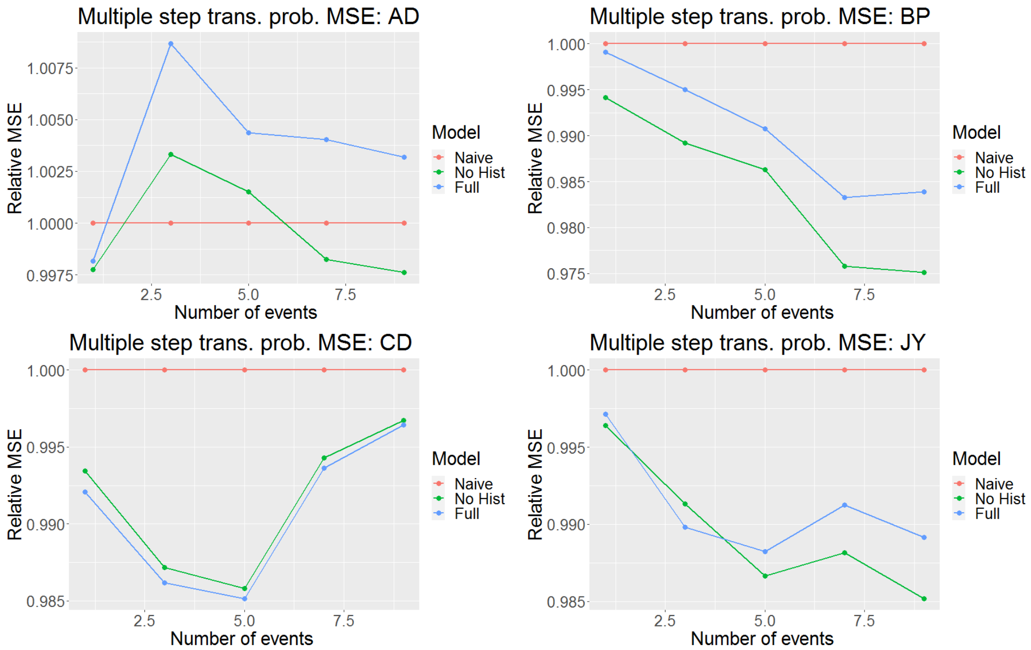


Figure 3.9: Relative MSE for the multiple step ask level transition probabilities, for the three models introduced in section 3.3.

3.4 Micro Price estimation

The model fitting procedure for the micro-price was conducted as follows. Firstly, the number of bins for the imbalance and spread were determined. Then, all market events in the training set were iterated over. For any of these market event times $\tilde{t}_n$, the subsequent time $\tilde{t}_{n+1}$ was also recorded. These times were in turn used to evaluate the touple $\{\mathcal{O}_{\tilde{t}_n}, \mathcal{O}_{\tilde{t}_{n+1}}\}$. From each pair of order books, the parameters $(I_{\tilde{t}_n}, S_{\tilde{t}_n}, I_{\tilde{t}_{n+1}}, S_{\tilde{t}_{n+1}}, dM)$ from section 2.5.1 were recorded. The states needed to be mirrored in order to ensure convergence of the micro-price - and each observation was therefore paired with a fictitious observation, characterized by $(1 - I_{\tilde{t}_n}, S_{\tilde{t}_n}, 1 - I_{\tilde{t}_{n+1}}, S_{\tilde{t}_{n+1}}, -dM)$. These values were then grouped and aggregated to form the frequency matrices $\tilde{R}$, $\tilde{T}$, $\tilde{Q}$.

After the training data had been iterated over, these frequency matrices were converted into empirical transition probability matrices $\hat{R}$, $\hat{T}$, $\hat{Q}$. This was done by dividing the number of observations in each element by the corresponding rowsum of the matrix $\hat{T} + \hat{Q}$. Then, the micro price adjustments was calculated according to the steps in 2.5.1. Figure 3.10 illustrates the micro-price adjustments x, expressed as fractions of a tick, for the four assets. In other words, the micro-price from (2.15) would be calculated as $z_t^{micro} = M_t + S_t \cdot x$.

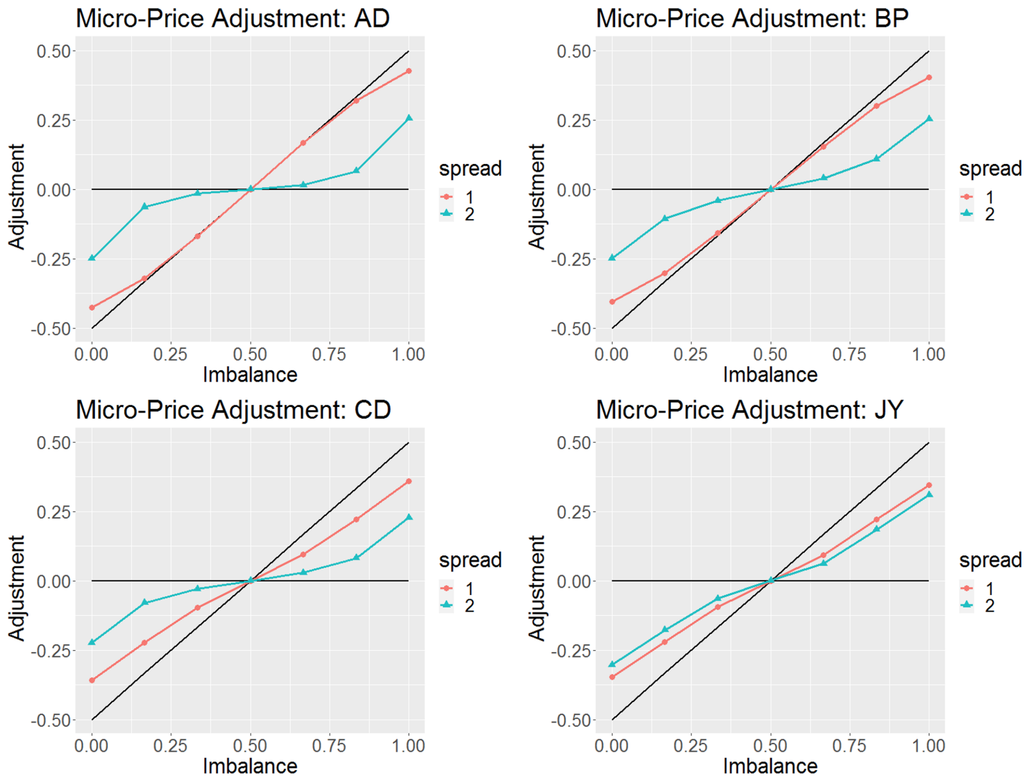


Figure 3.10: Micro price adjustment, expressed in number of ticks, for the four assets. The instances where $S_t > 2$ have been neglected.

The horizontal black line is the adjustment associated with the mid price M. This is always 0 since no adjustment needs to be made. The diagonal black line is the adjustment associated with the weighted mid price W. The weighted mid price increases linearly with the imbalance - and coincides with the mid price for an imbalance $I = 0$. Note that the micro-price tends to fall between the two - and that the adjustment for a spread of one tends to follow the weighted mid price more closely than the adjustment for a spread of two. The micro-price adjustments for $S > 2$ have been excluded from the figure.

3.5 MSE for the fair asset value

To evaluate the fair asset value models from section 2.5, the data was segmented into intervals of length $I = 3\,\text{min}$. This time was chosen to give a good balance between the estimates variance, which grows with time - and order book correlation, which decreases with time. For each time t_n the previous market event time $\underline{t_n}$ was used to evaluate the fair asset value estimates M_t, W_t, z_t^{micro}.

The order book at time $\underline{\tau_n} = \underline{t_{n+1}}$ was then used to evaluate another estimate of the fair asset value. Residuals for all combinations of fair value estimates were recorded as the difference between the model predictions - and the residuals were then used to calculate MSEs. The results can be observed in figure 3.11, where the double micro-price model is used as a benchmark.

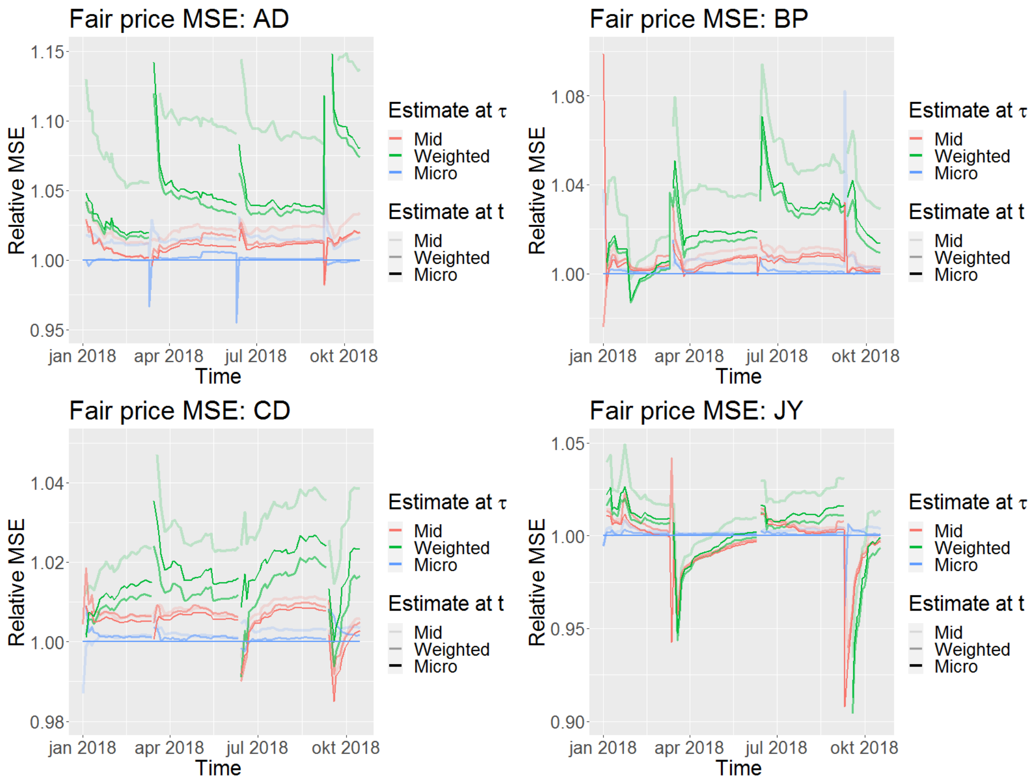


Figure 3.11: Relative MSE for the fair price estimates discussed in section 2.5. The transparency of the lines indicates what model was used to estimate the price at time t. The colour indicates what model was used to approximate the asset value at time τ.

The transparency of the lines indicates what model was used to estimate the asset value at time t. The colour indicates what model was used to approximate the asset value at time τ. Note that for the majority of observations, using the micro price model at both time points tends to yield the lowest MSE. The only exception is JY, where the micro-price results in a higher MSE for two of the time intervals.

3.6 Compound model evaluation

This section describes the evaluation procedure after the fitted Markov models and frequency models had been combined to form compound models. A generic illustration of the evaluation procedures can be seen in figure 3.12.

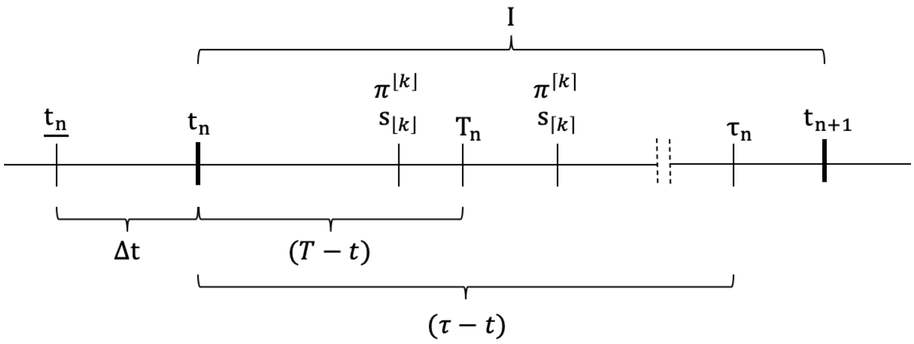


Figure 3.12: General evaluation procedure for the compound models. The time points $\{t_1, \ldots, t_{N-1}\}$ were separated by a fixed interval $I = 3\,\text{min}$. Each of these time points t_n represents the time of placement for a fictitious limit order. The associated time T_n represents the time that the order is filled or killed. $s_{\lfloor k \rfloor}$ and $s_{\lceil k \rceil}$ represent the previous and preceding market events with respect to T_n. k is the predicted number of orders according to the given market activity model.

3.6.1 MSE for ask level transition probabilities over time intervals

Evaluating the MSE for multiple step ask level transition probabilities, as described in section 3.3.1, yields some insight into the predictability of the model. However, the more natural investigation is how accurately the model predicts transaction probabilities over some time interval [t,T]. Therefore, the compound model was used to predict the transition probability within fixed time intervals.

Firstly, the number of market events was predicted. Secondly, the transition probabilities were evaluated using the Markov models. Predictions on the number of orders in the interval $T - t$ were done in the same way as in section 3.2.1. However, the time T_n in figure 3.12 is unlikely to coincide with one the predicted market events. This means that the expected number of events in the interval $[t_n,\, T_n]$ is unlikely to be an integer. Therefore, transition probabilities $p_{\lfloor k \rfloor}$ and $p_{\lceil k \rceil}$ at times $s_{\lfloor k \rfloor}$ and $s_{\lceil k \rceil}$ were evaluated. This was done by means of the transition matrices $\Pi^{\lfloor k \rfloor}$ and $\Pi^{\lceil k \rceil}$. Linear interpolation was then used to derive the final estimate of the transaction probability according to

$$p_k = p_{\lfloor k \rfloor} + (p_{\lceil k \rceil} - p_{\lfloor k \rfloor}) \cdot (\lceil k \rceil - k) \tag{3.3}$$

These transition probabilities were then compared to whether or not the ask level actually changed during the interval. This was quantified as $|f(z^a_{t_i}, z^a_{t_{i-1}})|$, where $f(x_t, x_{t-1})$ is defined as in (2.24). The residual between the two quantities was used to calculate the MSE. Figure 3.13 shows relative MSE plots of ask level transition probabilities, during intervals of length 100 ms. The fixed frequency model with a constant transition probability was used as a benchmark. Note that there are no clear indications of any model outperforming the others.

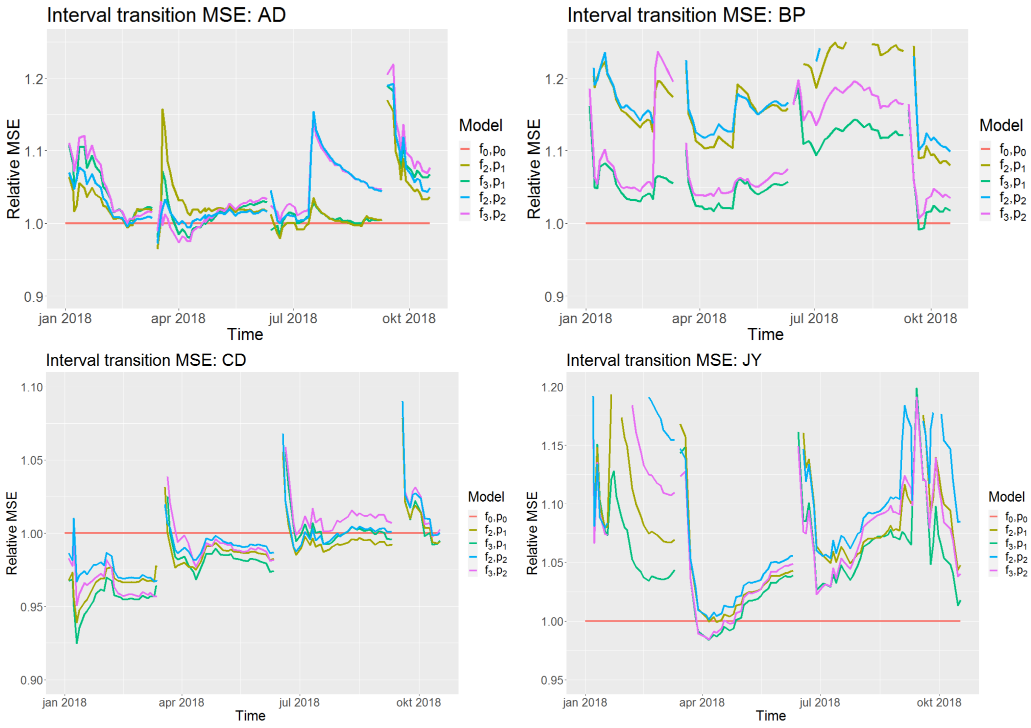


Figure 3.13: Relative MSE for the ask level transition probabilities, over time intervals of length 100ms, as a function of time. The MSE values are reset at the change of each contract, resulting in the three discontinuities across the time period.

3.6.2 MSE for cost estimates

During the evaluation procedure of the cost estimates, the data was partitioned into intervals of length $I = \tau - t = 3\,\text{min}$. The cost predictions were calculated at time t according to (2.19). This included evaluating the transition probabilities in the interval $[t_n,\, T_n]$, as well as evaluating the fair value of the asset. The asset price after a down move was always assumed to be the mid price.

Next, the ask level value was recorded at time T_n. This gave information of whether or not the order was filled. Lastly, the fair price of the asset was approximated at time τ_n. This was solely done using the micro-price, motivated by the results of figure 3.11. The residual was then calculated as the difference between the predicted cost and $z^a_{t_n} - z^{micro}_{T_n}$. The results for a latency $T - t = 100\,\text{ms}$ can be seen in figure 3.14, where a naive cost model of half the spread is used as a benchmark. Note that most of the models tend to give lower MSEs than the Naive model. Also note that the models making use of the micro-price tend to have the lowest MSEs.

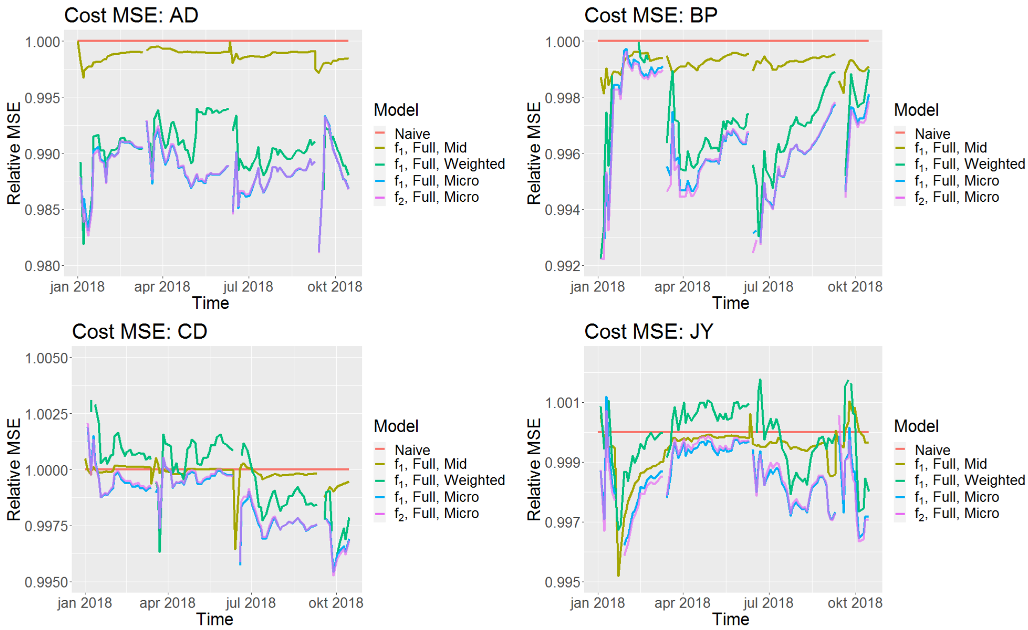


Figure 3.14: Relative MSE for the cost of a fill or kill buy limit order, where the naive cost model is used as a benchmark. Note that the models which make use of the micro-price have the lowest MSEs.

4

Discussion and Conclusion

This chapter discusses the results and possible improvements of the analysis conducted in chapter 3. The report is finished by a presentation of the final conclusions.

4.1 Frequency modelling

The existence of outliers were noted in figure 3.2, close to 10:00. Moreover, these outliers tend to have values that are significantly higher than what would be expected. A likely explanation for this phenomenon is the releasing of financial reports, which tend to spark trading activity. This effect could further be taken into account by allowing the function $g(\nu)$ to take on separate values at these specific points.

The fitted function $\hat{h}(\Delta t)$ in figure 3.4 does a decent job of flitting to market data. However, it can be noted that the function has a slightly sharper bend than the data, causing the residuals to be positive around 80 ms. Moreover, it seems like there is still some long term decay which the model does not capture. This can be noted since the residuals tend to be larger close to 1000 ms than they are by 500 ms. Exponentially decaying models were also evaluated, but they were unsuccessful in capturing the behaviour during the first 100 ms following a market event.

The fact that f_2 in figure 3.2.1 consistently maintains the lowest MSE out of the evaluated frequency models is interesting. This is because f_3 was designed to take the frequency shift in the interval [t,T] into account. An explanation may lie in the way $H(\Delta t, T - t)$ was evaluated. The underlying assumption of this derivation, which can be found in section A.2, was that the market event frequency was deterministic. In fact, the actual frequency follows some distribution. By taking this distribution into account, one may arrive at a better approximation. In further research, it would be especially interesting to take a Bayesian approach, much like in section 2.9.

One could discuss whether the relationship between the functions g and h is additive or multiplicative. A multiplicative relationship would indicate that all market participants act in a similar way. However, this is not necessarily the case. Market makers and algorithmic traders are usually active around the clock - and they are likely the fastest participants in the market. Hence, the market activity for small values of Δt may not scale linearly with the overall activity in the market. It would be interesting to investigate the market activity for small values of Δt in greater detail - and include models of market maker behaviour in the analysis.

4.2 State modelling

In section 2.6, it was assumed that the filtration $\mathcal{F}_t$ is well approximated by $\mathcal{F}_t^{\mathcal{O}}$. This is equivalent to assuming that all predictive information about the process can be attained from looking at historical order books. This is a natural assumption in the current setting, yet it is simultaneously neglects the effects that other factors might have on the process. Examples of these are the more fundamental indicators that value investors make use of to drive trades. However, these indicators are often used to determine the long term drift of the asset price. During the short time intervals associated with this kind of algorithmic trading, the long term drift of the process is assumed to be 0 - and these predictors become less informative. It would be interesting to further introduce the input of data external to the order book. There is for example active research on the impact that social media and news has on financial markets. [10] [11]

It can be noted that the one step transition probabilities generate better out of sample returns if the previous market event type is taken into account, as illustrated by figures 3.8 and 3.9. However, as can be seen in figure 3.13, the power of this predictability is reduced when a fixed time interval is considered. This is due to the added uncertainty of approximating the number of market events in the interval. With an improved model of market activity, the previous market event type may offer a significant improvement in predictability compared to the no history model.

4.3 Fair price estimation

Figure 3.11 indicates that using the micro-price to evaluate the fair price is preferable compared to using the mid price or weighted mid price. The impact of this is also notable in figure 3.14, where the micro-price models generate the lowest MSEs. This is not surprising since figure A.1 indicates that we are dealing with large tick assets. It would be interesting to also account for higher order data in the micro price evaluation to see if these results can be further improved upon.

In section 3.6.2 it was mentioned that the mid price was used as an estimate of the fair price in the case of a down move in the ask level. This simplifying assumption was made since there is no apparent information about the order book after the ask level has been conditioned to to change. A more detailed analysis could be performed to investigate the likelihoods of different market event books after a ask level transition has been observed. Note however that, since the transition probability tends to be small, this term has limited importance on the cost function (2.19).

4.4 Conclusions

The results show that market events tend to be closely followed by new ones. The type of the previous market event yields some predictability over the subsequent order book evolution. However, the marginal utility of including this information in the cost function is small compared to that of improving the fair value estimate of the asset. For this, the micro-price is found to offer a substantial improvement over the mid price, as well as the weighted mid-price. The significant improvement in cost modeling can be capitalized on to construct trading strategies for futures contracts in currency markets.

In further research, it would be interesting to explore how different models of market maker behaviour can be used to infer the observations of this paper. It would also be interesting to investigate the time it takes for the market to respond to news events - like the ones which seem to have caused the increase in market activity around 10:00. One could attempt to couple the level of optimism in these news statements with the ask level transition probabilities to see if it yields some degree of predictive power. The time between measurements was selected to be 3 min. A rigorous examination of the best delay could be performed - potentially resulting in better signal to noise ratios. It would also be insightful to penalize the event of non-execution. Lastly, it would be interesting to relax the assumption of fill or kill orders to include general limit orders. However, anonymity of the markets make it hard to tell whether subsequent orders are market responses or simply part of the initial traders strategy. Therefore, it may be challenging to simply utilize historical order book data in such an investigation.

Bibliography

[1] Alain P Chaboud, Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega. Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance*, 69(5):2045–2084, 2014.

[2] T. Björk. *Arbitrage Theory in Continuous Time*. Oxford Finance Series. Oxford University Press, Incorporated, 2004.

[3] Rajeshwari Majumdar, Phanuel Mariano, Lowen Peng, and Anthony Sisti. A derivation of the black-scholes option pricing model using a central limit theorem argument, 2018.

[4] Yacine Aït-Sahalia, Per A Mykland, and Lan Zhang. Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, 160(1):160–175, 2011.

[5] Anirban Chakraborti, Ioane Muni Toke, Marco Patriarca, and Frédéric Abergel. Econophysics review: I. empirical facts. *Quantitative Finance*, 11(7):991–1012, 2011.

[6] Khalil Dayri and Mathieu Rosenbaum. Large tick assets: implicit spread and optimal tick size, 2013.

[7] C. Moallemi and Kai Yuan. A model for queue position valuation in a limit order book. 2016.

[8] Sasha Stoikov. The micro-price: A high frequency estimator of future prices. *Available at SSRN 2970694*, 2017.

[9] Christopher C. Strelioff, James P. Crutchfield, and Alfred W. Hübler. Inferring markov chains: Bayesian estimation, model comparison, entropy rate, and out-of-class modeling. *Physical Review E*, 76(1), Jul 2007.

[10] B. Shravan Kumar and Vadlamani Ravi. A survey of the applications of text mining in financial domain. *Knowledge-Based Systems*, 114:128 – 147, 2016.

[11] Hamed Naderi Semiromi, Stefan Lessmann, and Wiebke Peters. News will tell: Forecasting foreign exchange rates based on news story events in the economy calendar. *The North American Journal of Economics and Finance*, 52:101181, 2020.

A

Appendix 1

A.1 Derivation of konstant k

The condition that $\mathbb{E}_{\mathcal{F}^{\mathcal{O}}_{t_i}}[g(u)] = 1$, where $u \sim U[0, 24)$ is equivalent to the condition $\int_0^{24} g(x)dx = 24$. Let $h(x) = a_1 f_{\mu_1,\gamma_1}(x) + a_2 f_{\mu_2,\gamma_2}(x) + a_3 f_{\mu_3,\gamma_3}(x)$. The primitive function $H(x)$ becomes (up to a constant)

$$\begin{aligned} H(x) &= a_1 F_{\mu_1,\gamma_1}(x) + a_2 F_{\mu_2,\gamma_2}(x) + a_3 F_{\mu_3,\gamma_3}(x) \\ F_{x_0,\gamma}(x) &= \frac{1}{\pi} \cdot \arctan(\frac{x - x_0}{\gamma}) + \frac{1}{2} \end{aligned} \tag{A.1}$$

One therefore arrives at the following expression for k

$$\begin{aligned} 24 &= \int_0^{24} g(x)dx \\ &= 23k + \int_0^{15} h(x)dx + (1-c)\int_{15}^{16} h(x)dx + (1-c)\int_{17}^{18} h(x)dx + \int_{18}^{24} h(x)dx \\ &= 23k + c[H(15) - H(18)] + (1-c)[H(16) - H(17)] + H(24) - H(0) \\ \Rightarrow k &= \frac{24}{23} - \frac{1}{23}\left(c[H(15) - H(18)] + (1-c)[H(16) - H(17)] + H(24) - H(0)\right) \end{aligned} \tag{A.2}$$

A.2 Average market activity during a time interval

Assume that $f_0 \cdot g(\nu) \cdot h(\Delta t)$ models all the variability in market activity. Moreover, assume that $1/f_0 \cdot g(\nu) \cdot h(\Delta t) > T - t$. In this case, one would not get any market events in the interval $[t, T]$. Therefore, no added compensation would be needed and the function $H(\nu, \Delta t, T - t)$ would equal $h(\Delta t)$. Stated mathematically

$$f_0 \cdot g(\nu) \cdot H(\nu, \Delta t, T - t) = f_0 \cdot g(\nu) \cdot h(\Delta t), \quad \text{if} \quad h(\Delta t) \leq 1/(T - t)g(\nu)f_0. \tag{A.3}$$

However, if $1/f_0 \cdot g(\nu) \cdot h(\Delta t) < T - t$, at least get one market event would occur in the interval $[t, T]$. Moreover, after this event had occurred, the market activity would be $f_0 \cdot g(\nu) \cdot h(0)$. The average market activity during the interval $[t, T]$ would therefore be

$$f_0 \cdot g(\nu) \cdot H(\Delta t, T - t) = f_0 \cdot g(\nu) \cdot h(\Delta t) \cdot \frac{\bar{t} - t}{T - t} + f_0 \cdot g(\nu) \cdot h(0) \cdot \frac{T - \bar{t}}{T - t}, \quad \text{(A.4)}$$

where $\bar{t}$ is the time of the market event preceding t, which (under the assumption of deterministic order frequency) can also be written on the form

$$\bar{t} = t + \frac{1}{f_0 \cdot g(\nu) \cdot h(\Delta t)}. \quad \text{(A.5)}$$

By simplifying the first two equations in this section one arrives at the following estimate for $H(\Delta t, T - t)$

$$H(\nu, \Delta t, T - t) = \begin{cases} h(\Delta t) \ , & \text{if} \quad h(\Delta t) \leq x \\ h(0) + x(1 - \frac{h(0)}{h(\Delta t)}) \ , & \text{if} \quad h(\Delta t) > x \\ x = 1/(T - t) g(\nu) f_0. \end{cases} \quad \text{(A.6)}$$

A.3 Spreads

The following graphs show histograms over the spread distributions for four assets: Austalian dollar (AD), Brittish pounds (BP), Canadian dollar (CD) and Japanese yen (JY). Note that they rarely exceed a spread of two ticks, which is why figure 3.10 only includes compensations for spreads of 1 and 2.

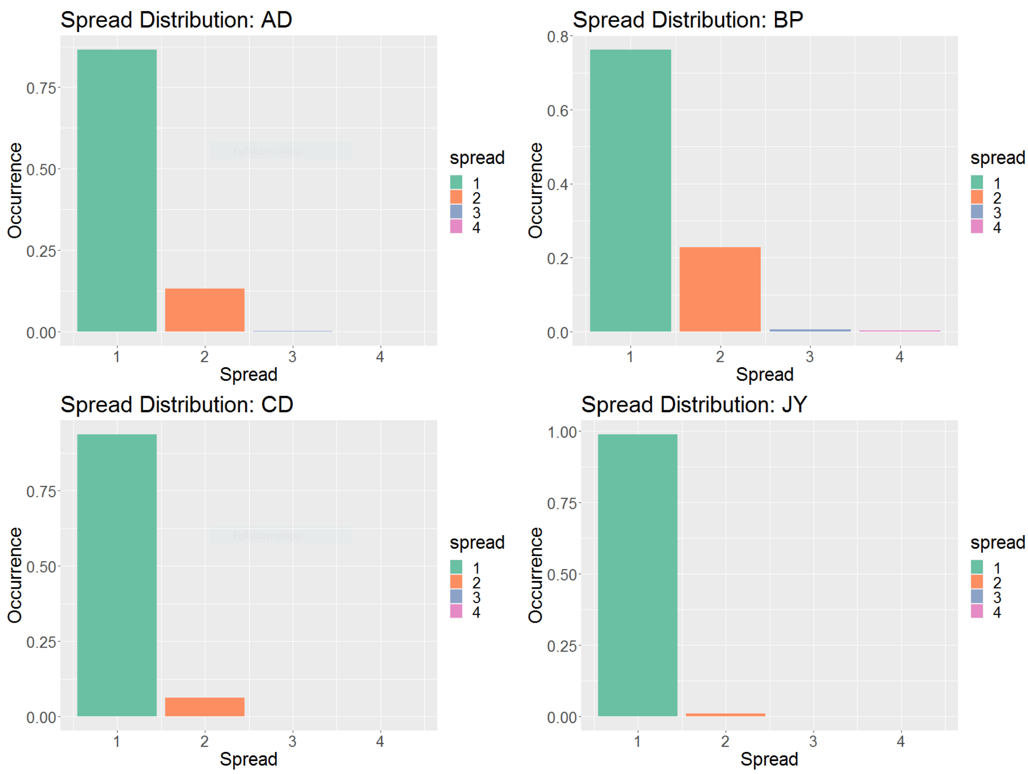


Figure A.1: Histograms for the spread distributions of the four assets. Note that BP is the most likely to have a spread of two ticks and that JY is the most likely to have a one tick spread.

A.4 Absolute MSE plots

Many of the MSE plots in chapter 3 displayed the relative MSE compared to some benchmark. This was done in order enhance differences between the models. The following section contains the corresponding MSE plots, without any adjustments.

Figure A.2 shows a comparison between the different frequency models, in comparing the number of market events during a time interval of 100 ms. The corresponding relative MSE plot can be found in figure 3.6.

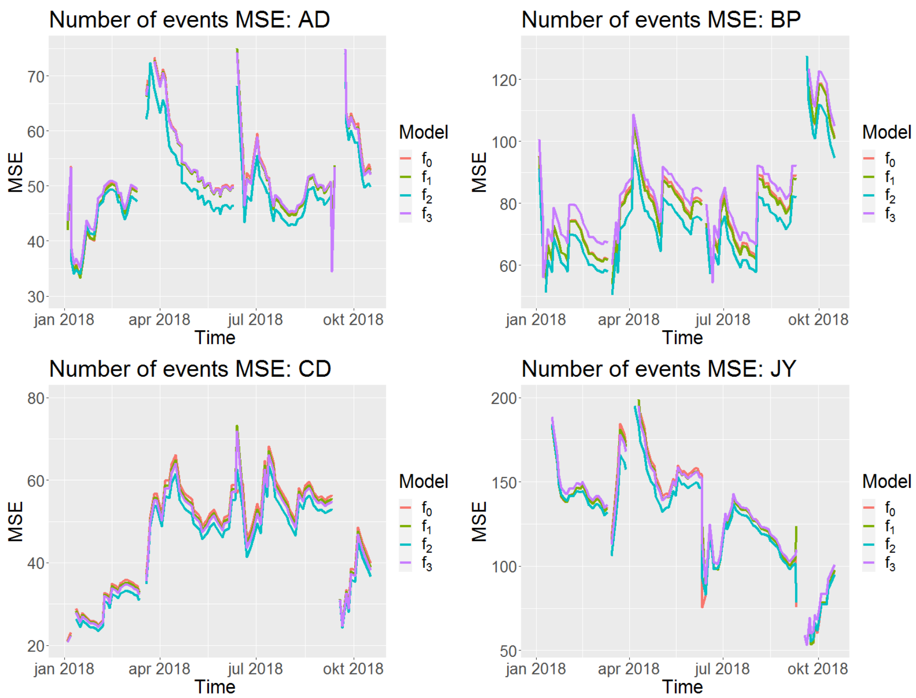


Figure A.2: Comparison between the frequency models, in comparing the number of market events during a 100 ms time interval.

Figure A.3 shows a comparison between the different markov models, described in section 3.3, in predicting the probability of a transition in the ask level at the time of the next market event. The corresponding relative MSE plot can be found in figure 3.8.

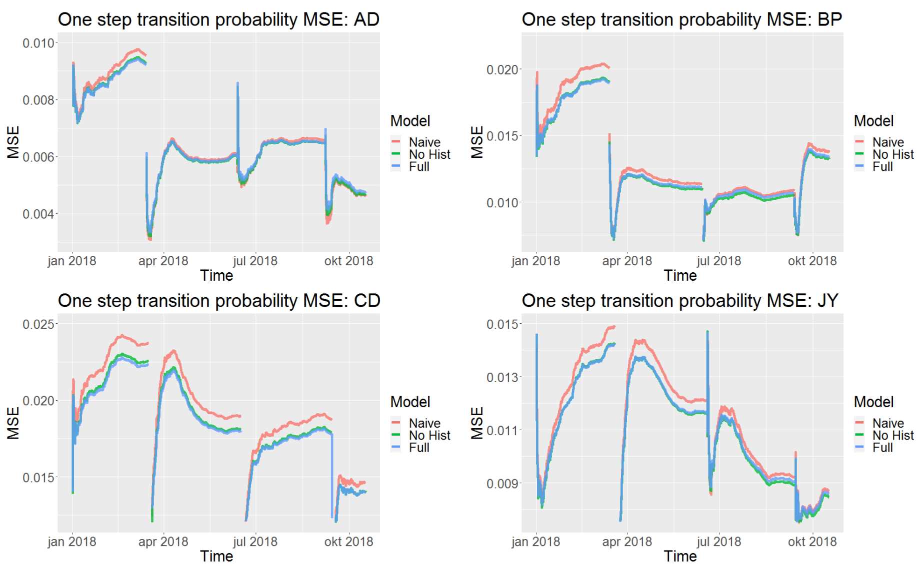


Figure A.3: Comparison between the markov models, described in section 3.3, in predicting the probability of a transition in the ask level at the time of the next market event.

Figure A.4 shows a comparison between the different markov models, described in section 3.3. The quantity of interest is the probability of a transition in the ask level, sometime during the following n market events. The corresponding relative MSE plot can be found in figure 3.9. Note that, unlike in figure A.3, all of the data associated with a particular number of market events n has been summarized into a single point. This reduces the complexity, but it also limits insights over how the markov models compare over time.

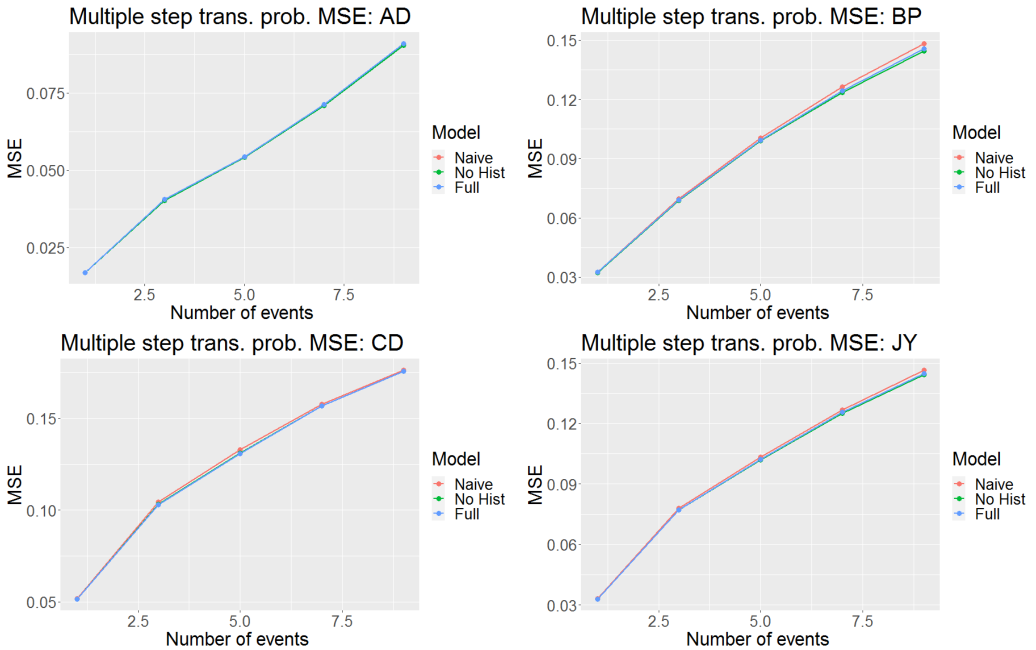


Figure A.4: Comparison between the markov models, described in section 3.3 on predicting the probability of a transition in the ask level - sometime during the following n market events. The x-axis shows the number of market events n. The y-axis shows the MSE for the models.

Figure A.5 shows a comparison between the compound models, in predicting the probability of a transition in the ask level, during a time interval of 100 ms. The benchmark model $\{f_0, p_0\}$ was highlighted in black for illustration purposes. Figure 3.13 shows the relative MSEs for some of the key models in figure A.5.

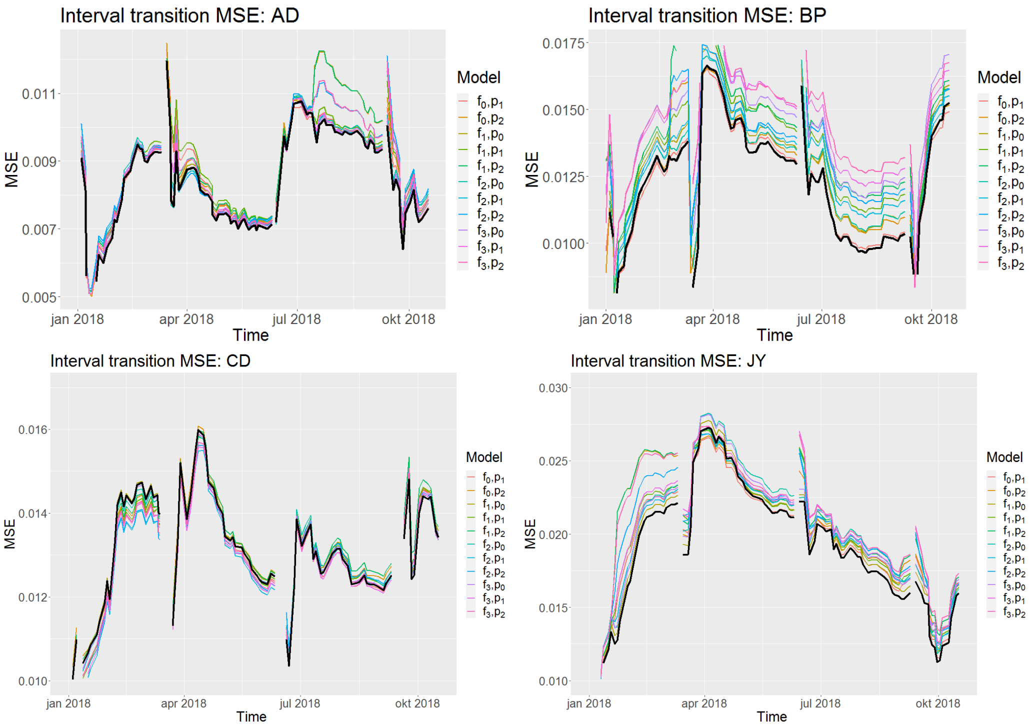


Figure A.5: Comparison between the compound models, in predicting the probability of a transition in the ask level, during a time interval of 100 ms. The benchmark model $\{f_0, p_0\}$ was highlighted in black for illustration purposes.

Figure A.6 shows a comparison between all combinations of fair price estimates. Figure 3.11 shows the associated relative MSEs, withe the $\{\text{Micro}, \text{Micro}\}$-model as a benchmark.

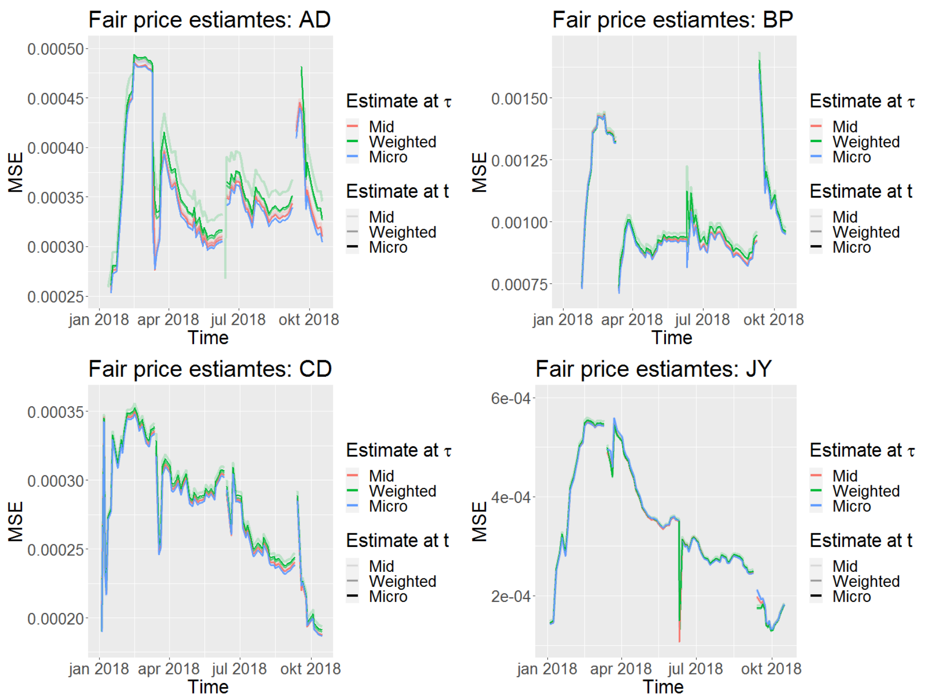


Figure A.6: comparison between all combinations of fair price estimates. The models were compared in terms of MSEs, which were evaluated according to section 3.5.

Figure A.7 shows a comparison between the different compound models in predicting the cost of a fill or kill limit order. Figure 3.14 shows the associated relative MSEs, withe the naive model as a benchmark.

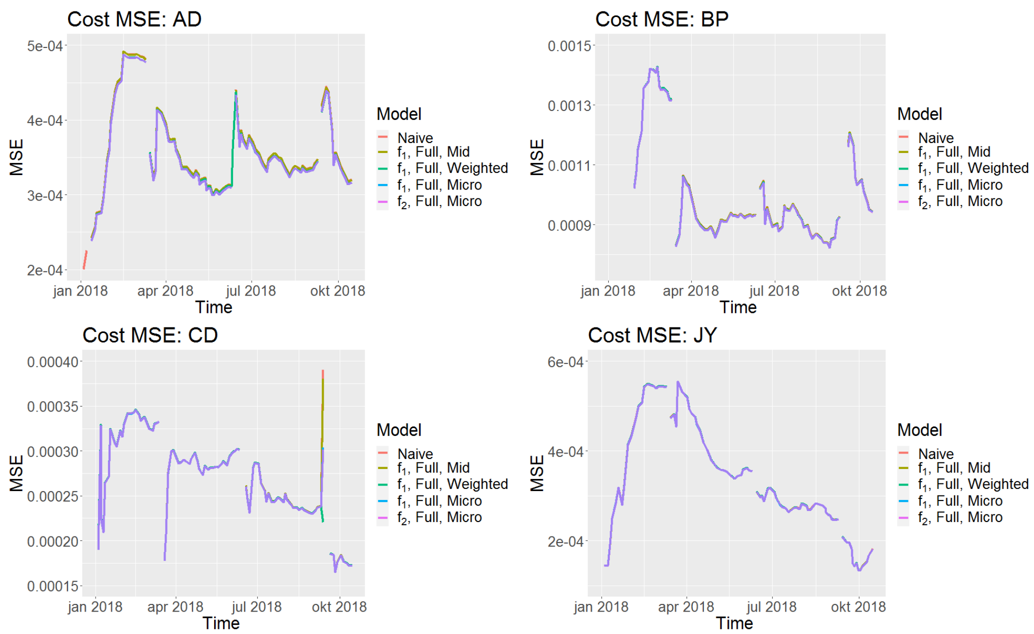


Figure A.7: Comparison between the different compound models in predicting the cost of a fill or kill limit order. The models were compared in terms of their MSEs, which were evaluated according to section 3.6.2.